Towards large deviations in stochastic systems with memory

by

Massimo Cavallaro

School of Mathematical Sciences
Queen Mary University of London

Submitted to the University of London in partial fulfilment of the requirements of the Degree of Doctor of Philosophy

August 2016

I, Massimo Cavallaro, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Signature:

Date: October 27, 2016

Details of collaboration and publications:

- M. Cavallaro, R. J. Mondragón, and R. J. Harris, Temporally correlated zero-range process with open boundaries: Steady state and fluctuations, Phys. Rev. E. 92(2):022137, 2015.

- M. Cavallaro and R. J. Harris, A framework for the direct evaluation of large deviations in non-Markovian processes, arXiv:1603.05634, to appear as a Letter in J. Phys. A: Math. Theor., 2016.

  M. C. carried out essentially all calculations and simulations and is lead author for the manuscripts.

Abstract

The theory of large deviations can help to shed light on systems in non-equilibrium statistical mechanics and, more generically, on non-reversible stochastic processes. For this purpose, we target trajectories in space–time rather than static configurations and study time-extensive observables. This suggests that the details of the evolution law—such as the presence of time correlations—take on a major role. In this thesis, we investigate selected models with stochastic dynamics that incorporate memory by means of different mechanisms, devise a numerical approach for such models, and quantify to what extent the memory affects the large deviation functionals. The results are relevant for real-world situations, where simplified memoryless (Markovian) models may not always be appropriate.

After an original introduction to the mathematics of stochastic processes, we explore, analytically and numerically, an open-boundary zero-range process which incorporates memory by means of hidden variables that affect particle congestion. We derive the exact solution for the steady state of the one-site system, as well as a mean-field approximation for larger one-dimensional lattices. Then, we focus on the large deviation properties of the particle current in such a system. This reveals that the time correlations can be apparently absorbed in a memoryless description for the steady state and the small fluctuation regime. However, they can dramatically alter the probability of rare currents. Different regimes are separated by dynamical phase transitions. Subsequently, we address systems in which the memory cannot be encoded in hidden variables or the waiting-time distributions depend on the whole trajectory. Here, the difficulty in obtaining exact analytical results is exacerbated. To tackle these systems, we have proposed a version of the so-called "cloning" algorithm for the evaluation of large deviations that can be applied consistently for both Markovian and non-Markovian dynamics. The efficacy of this approach is confirmed by numerical results for some of the rare non-Markovian models whose large deviation functions can be obtained exactly. We finally adapt this machinery to a technological problem, specifically the performance evaluation of communication systems, where temporal correlations and large deviations are important.

Table of Contents

List of Figures

Acknowledgements

I wish to express my warmest gratitude to my supervisor Rosemary Harris and co-supervisor Raúl Mondragón for their continued support and close guidance. I believe that this research would not have been possible without their thoughtful insight, challenging questions, and wise suggestions. I doubt that I will ever be able to express my full gratitude to Rosemary, whose expertise and patience have been invaluable.

I am very grateful to all the friends I made at the School of Mathematical Sciences for providing some very enjoyable time, as well as sharing some annoying aspects of the research experience. I feel that André Nock, Andrea Cairoli, Mike Rigby, Moreno Bonaventura, Nils Haug, and Rodrigo Sanchez deserve to be mentioned here as my research certainly benefited to some extent from their casual but bright suggestions and comments. Thanks also to Enzo Nicosia, Oscar Bandtlow, and Vito Latora who never denied support whenever I asked. Additional thanks go to Hugo Touchette, Pablo Hurtado, and Stefan Grosskinsky with whom I have had the pleasure of discussing part of this work, and Ludger Santen for granting hospitality at the Universität des Saarlandes.

Above all, it is very likely that I would not have been able to submit this specialised dissertation on history/age-dependent events without the support that my family provided throughout my entire life. I am especially grateful to my partner and best friend Mariangela whose love truly made a difference in my life.

The research utilised Queen Mary's MidPlus computational facilities, supported by QMUL Research-IT and funded by EPSRC Grant No. EP/K000128/1.

Introduction

Contents

Scientists have long been tempted to imagine that there is a way to represent a real-world system in terms of a sheer number of variables, whose states at earlier times uniquely determine their future states [8]. This view is in fact well accepted if we exclude the realm of Quantum Physics. However, a complete knowledge of at least some such variables is inaccessible in practice. Stochastic modelling is the selected way to cope with such a lack of information; the general strategy is to approximate the influence of the inaccessible information on future states simply with random noise. This thesis deals with three important aspects of stochastic modelling, namely non-equilibrium dynamics, large deviations, and memory, and is motivated in this introduction.

The importance of being unbalanced and rare

As a first step, rather than attempting to describe the state of all the single variables one by one, we can represent the system as a smaller collection of *random variables*, i.e., functionals that can take on a set of possible different values (each with an associated probability). In other words, the state of the system at time t is represented by a probability distribution for the configurations that the system can adopt, which we represent by a column vector $|P(t)\rangle$. We still wish to incorporate causality in a stochastic process. Hence, the next step is to write an evolution law of the form

$$|P(t)\rangle = \mathbf{T}(t - t_0)|P(t_0)\rangle,$$

where $\mathbf{T}(t-t_0)$ is an operator that encodes for the dynamics and can depend on the initial time t_0, as well as on the duration $t-t_0$ of the evolution time, of course. We identify two important classes of stochastic dynamics: those satisfying the Markov property and those that do not. Roughly speaking, in a Markov process it is possible to make predictions for the future behaviour based only on the present, and such predictions are as accurate as those based on the knowledge of the past history.

As long as we focus on the typical behaviour of a system (that is, the values of the most probable realisations of its random variables) some of the details of the evolution law can be neglected. This approach is often appropriate when we deal with very large systems, where the probability of a typical configuration overwhelms that of atypical ones. However, there are situations in which the behaviour of interest is not typical, but rather *rare* and *atypical*. For example, the transport of energy, particles, or vehicles could be enhanced by exceptional coherent configurations, or occasionally delayed when an instantaneous situation that will be later referred to as congestion or condensation occurs [29]. Similarly, in communication or transportation networks it is very important to predict how likely is to have interruptions or packet loss [31]. In climate science the assessment of the likelihood of extreme weather is of central importance, even compared to the prediction of the average global trend [4, 10]. Another reason for the interest in rare events is that small observation scales are now accessible to some experiments, as in molecular physics [5, 26]. Moreover, rare events help to shed light on the foundations of non-equilibrium statistical mechanics, just as they play an important role in defining the thermodynamic potentials in equilibrium Statistical Mechanics [34]. Specifically, our focus is on rare values of time-extensive random observables, whose realisations are determined by the details of the evolution law.

One of the major long-standing problems in Science is to characterise and predict the behaviour of a macroscopic system, given its microscopic model. Statistical Mechanics, as designated by the work of W. J. Gibbs [25], has been successfully applied to a specific class of systems, whose stochastic dynamics obey a specific symmetry, known as the *detailed balance* and are said, in the physics literature, to be at *equilibrium*. Roughly speaking, the detailed balance requires that the number of transitions from a microscopic state to another one, is equal to the number of those in the opposite direction. As we will see later, this condition is strong but useful, as it greatly simplifies the analysis of a stochastic process, when satisfied. When that is not the case, i.e., the state is *unbalanced*, the nett number of transitions defines a non-vanishing probability current, which in turns can represent a physical observable such as the energy transfer or the particle drift (quantities which are often assumed to be time-extensive, i.e., to grow linearly with time). We can hardly overestimate the importance of currents for applications, as

they are naturally required to model transport, growth, damage, production, or any other "activity" in a system, and are found across many applied disciplines, ranging from Physics (e.g, rheology, quantum transport, molecular motors), engineering (vehicular traffic, electronic engineering, telecommunications), biology (chemotaxis, population dynamics). Modern accounts of these topics are given, e.g., in the books of Mahnke et al. [21] and Schadschneider et al. [29].

A distribution $|P(t)\rangle$ that does not vary with time, whilst permitting non-zero currents, is said to represent a *non-equilibrium stationary state* (NESS). Additionally, it is often said that such currents characterise the NESS [36]. It is important to stress that, to treat systems far from equilibrium, it is necessary to consider explicitly the microscopic dynamics as NESSs do not necessary have an analytical expression as simple as the detailed-balance one. As a consequence, analytical progress is challenging and inspires a research branch on exactly solvable models, which has the aim to provide some insight into the foundations of physics [30]. While a general framework for the characterisation of systems far from the thermal equilibrium is at a primitive stage, nowadays large deviation theory plays a central role for building a comprehensive theory of these systems [34]. This is the selected tool that allows us to evaluate the fluctuations of time-extensive observables such as the currents.

Non-Markovian modelling

A better strategy to cope with the lack of information on the present state, is to supplement the dynamical rules for the time-evolution of $|P(t)\rangle$ with a dependence on some information on the past events (which can be regarded as memory). Such a strategy adapts well to the realistic situation where we have an incomplete instantaneous description of the real-world system we wish to model, but we learned, from observations or theory, the events that yield configurations changes at different instants are temporally correlated [2, 4, 15, 17, 22–24, 27, 32, 33]. As shown throughout the thesis, this is the essence of non-Markovian modelling.

Despite the fact that dealing with temporal correlations is often both useful and appealing, non-equilibrium systems with memory have been investigated more rarely than those without memory, due to limited analytical techniques for analysing non-Markovian processes. We mention that, recently, some research focused on this niche: a necessarily incomplete list of publications is, e.g., [1, 3, 7, 9, 11–14, 16, 19, 20, 28, 35], see also the references therein.

Our work intends to contribute in the same direction by analysing selected stochastic models of non-equilibrium systems and developing numerical tools to deal with memory in these systems. Along this thesis we demonstrate, both analytically and numerically, that the fluctuations of time-extensive quantities *cannot* be predicted solely on the basis of a Markovian description. The thesis is structured as follows. In chapter 1 we discuss several aspects of stochastic processes, covering Markovian and non-Markovian stochastic dynamics, equilibrium and non-equilibrium stationary states, the large deviation approach to non-equilibrium systems. We also introduce important algorithms to simulate the stochastic dynamics. This chapter is organised in an original way to demonstrate interesting links between the mathematics and physics literature and show some consequences of the different ways to deal with time and temporal correlations in stochastic processes.

Chapters 2 and 3 are based on Cavallaro et al. [7] and are mainly concerned with a specific non-Markovian stochastic model of particles on lattice. Such a model incorporates memory by means of "hidden" phases and is referred to as the *on-off zero-range process* (on-off ZRP). After a general introduction to the interacting-particle systems (IPSs) and their quantum Hamiltonian formulation, the on-off ZRP and its NESS are analysed in chapter 2. Specifically, we derive its exact analytical expression for the single-site lattice and a mean-field approximation for arbitrary lattice topology. In chapter 3 the large deviation properties of the particle current for the on-off ZRP are examined. We show that, although the particle distribution is well described by an effective Markovian solution, with a structure similar to the NESS, the probability of rare currents differs from the memory-less case. In particular, we find evidence for memory-induced dynamical phase transitions. The findings of this chapter are supported by numerical results, obtained using the "cloning" algorithm of Lecomte and Tailleur [18], which has been conceived for continuous-time Markov processes. In fact, this numerical scheme is applicable to the on-off ZRP only because such a model possesses a Markov representation in terms of hidden variables, while some classes of non-Markovian processes do not benefit of such a representation.

To tackle the remaining classes of non-Markovian processes, we therefore devise a more general algorithm. This is presented in chapter 4, along with exhaustive tests that demonstrate the validity of the method using non-Markovian variants of an ion-channel model and the *totally asymmetric exclusion process*, recovering results obtainable by other means. General accuracy and performance issues of the cloning methods are also considered. This chapter is based on the work published in Cavallaro and Harris [6].

Finally, chapter 5 is concerned with a real-world application of the cloning method of chapter 4 to Queueing Theory. This exploits the analogies between the non-equilibrium

physics of the IPSs and the technology of packet delivery in teletraffic engineering. In fact, in both cases, non-Markovian modelling is appropriate and rare events are of main concern. We demonstrate that the cloning algorithm of chapter 4 is suitable for the analysis of packet traffic. The content of this chapter is intended to be expanded and prepared for publication.

The thesis is concluded in chapter 6, where we outline some questions left open during the research process and suggest some ideas that may further contribute to comprehend memory in the framework of large deviations.

Bibliography

[1] T. Brandes and C. Emary. Feedback control of waiting times. *Phys. Rev. E*, 93(4): 042103, 2016.

[2] D. Bratsun, D. Volfson, L. S. Tsimring, and J. Hasty. Delay-induced stochastic oscillations in gene regulation. *Proc. Natl. Acad. Sci. USA*, 102(41):14593–8, 2005.

[3] T. Brett and T. Galla. Stochastic processes with distributed delays: Chemical langevin equation and linear-noise approximation. *Phys. Rev. Lett.*, 110(25):250601, 2013.

[4] A. Bunde, J. F. Eichner, J. W. Kantelhardt, and S. Havlin. Long-term memory: A natural mechanism for the clustering of extreme events and anomalous residual times in climate records. *Phys. Rev. Lett.*, 94(4):048701, 2005.

[5] C. Bustamante, J. Liphardt, and F. Ritort. The nonequilibrium thermodynamics of small systems. *Phys. Today*, 58(7):43, 2005.

[6] M. Cavallaro and R. J. Harris. A framework for the direct evaluation of large deviations in non-Markovian processes. 2016. To appear as a Letter in J. Phys. A: Math. Theor.

[7] M. Cavallaro, R. J. Mondragón, and R. J. Harris. Temporally correlated zero-range process with open boundaries: Steady state and fluctuations. *Phys. Rev. E*, 92(2): 022137, 2015.

[8] C. Cercignani. *Ludwig Boltzmann: The Man Who Trusted Atoms*. Oxford University Press, Oxford UK, 1998.

[9] R. J. Concannon and R. A. Blythe. Spatiotemporally complete condensation in a non-Poissonian exclusion process. *Phys. Rev. Lett.*, 112(5):050603, 2014.

[10] C. B. Field, Barros V., T. F. Stocker, and Q. Dahe. *Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, New York NY, 2012.

[11] R. García-García. Nonadiabatic entropy production for non-Markov dynamics. 86

(3):031117, 2012.

[12] C. D. Greenman and T. Chou. Kinetic theory of age-structured stochastic birth-death processes. *Phys. Rev. E*, 93(1):012112, 2016.

[13] R. J. Harris. Fluctuations in interacting particle systems with memory. *J. Stat. Mech.*, 2015(7):P07021, 2015.

[14] R. J. Harris. Random walkers with extreme value memory: modelling the peak-end rule. *New J. Phys.*, 17(5):053049, 2015.

[15] M. H. Jensen, K. Sneppen, and G. Tiana. Sustained oscillations and time delays in gene expression of protein Hes1. *FEBS Letters*, 541(1-3):176–177, 2003.

[16] D. Khoromskaia, R. J. Harris, and S. Grosskinsky. Dynamics of non-Markovian exclusion processes. *J. Stat. Mech.*, 2014(12):P12013, 2014.

[17] L. F. Lafuerza and R. Toral. Stochastic description of delayed systems. *Phil. Trans. R. Soc. A*, 371(1999):20120458–20120458, 2013.

[18] V. Lecomte and J. Tailleur. A numerical approach to large deviations in continuous time. *J. Stat. Mech.*, 2007(03):P03004–P03004, 2007.

[19] Y. T. Lin and T. Galla. Bursting noise in gene expression dynamics: linking microscopic and mesoscopic models. *J. R. Soc. Interface*, 13(114):20150772, 2016.

[20] C. Maes, S. Safaverdi, P. Visco, and F. van Wijland. Fluctuation-response relations for nonequilibrium diffusions with memory. *Phys. Rev. E*, 87(2):022125, 2013.

[21] R. Mahnke, J. Kaupuzs, and I. Lubashevsky. *Physics of Stochastic Processes: How Randomness Acts in Time*. Wiley, Weinheim Germany, 2009.

[22] R. N. Mantegna and H. E. Stanley. *Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge University Press, Cambridge UK, 1999.

[23] J. Mikisz, J. Poleszczuk, M. Bodnar, and U. Foryś. Stochastic models of gene expression with delayed degradation. *Bull. Math. Biol.*, 73(9):2231–2247, 2011.

[24] J. M. Pedraza and J. Paulsson. Effects of molecular memory and bursting on fluctuations in gene expression. *Science*, 319(5861):339–43, 2008.

[25] L. E. Reichl. *A Modern Course in Statistical Physics*. Wiley, New York NY, 2009.

[26] F. Ritort. Nonequilibrium fluctuations in small systems: From physics to biology. In *Advances in Chemical Physics*, volume 137, pages 31–123. John Wiley & Sons, Hoboken NJ, 2008.

[27] M. R Roussel and R. Zhu. Validation of an algorithm for delay stochastic simulation of transcription and translation in prokaryotic gene expression. *Phys. Biol.*, 3(4): 274–284, 2006.

[28] K. J. Rubin, K. Lawler, P. Sollich, and T. Ng. Memory effects in biochemical networks as the natural counterpart of extrinsic noise. *J. Theo. Biol.*, 357:245–267, 2014.

[29] A. Schadschneider, D. Chowdhury, and K. Nishinari. *Stochastic Transport in Complex Systems: From Molecules to Vehicles*. Elsevier, Oxford UK, 2010.

[30] G. M. Schütz. Exactly solvable models for many-body systems far from equilibrium. *Phase transitions and critical phenomena*, 19:1–251, 2001.

[31] R. D. Smith. The dynamics of internet traffic: self-similarity, self-organization, and complex phenomena. *Adv. Complex Syst.*, 14(06):905–949, 2011.

[32] D. Sornette and A. Helmstetter. Endogenous versus exogenous shocks in systems with memory. *Physica A*, 318(3-4):577–591, 2003.

[33] J. Stehlé, A. Barrat, and G. Bianconi. Dynamical and bursty interactions in social networks. *Phys. Rev. E*, 81(3):035101, 2010.

[34] H. Touchette. The large deviation approach to statistical mechanics. *Phys. Rep.*, 478(1-3):1–69, 2009.

[35] P. Van Mieghem and R. van de Bovenkamp. Non-Markovian infection spread dramatically alters the susceptible-infected-susceptible epidemic threshold in networks. *Phys. Rev. Lett.*, 110(10):108701, 2013.

[36] R. K. P. Zia and B. Schmittmann. Probability currents as principal characteristics in the statistical mechanics of non-equilibrium steady states. *J. Stat. Mech.*, 2007 (07):P07012–P07012, 2007.

1 Stochastic processes

Contents

We motivated the importance of including randomness in models of real-world systems in the previous chapter. In this chapter, we introduce the theory of stochastic processes, which allows us to describe rigorously such models. A stochastic process is defined as an ordered collection of random variables with values in $\mathcal{S}$, indexed by some set $\mathcal{T}$ [34]. Hereafter, we will implicitly consider $\mathcal{S}$ to be a discrete set, although most of the results remains valid if $\mathcal{S}$ is continuous. On the other hand, $\mathcal{T}$ can either be the set of natural numbers, when we say that the process evolves in *discrete time*, or $\mathcal{T} = [0, \infty)$, when the process is referred to as in *continuous time*. In fact, we think of $\mathcal{T}$ as the set of some instants in time while the stochastic process represents the time-evolution of a

system. A specific realization of a stochastic process is called a *sample path*, *trajectory* or *history*.

A typical choice is to assume that, conditional on its value at an instant in time, the future value of a random variable does not depend on the previous ones. This case is considered in sections 1.1 and 1.2, which deal with discrete-time and continuous-time Markov processes, respectively. In sections 1.3 and 1.5 we relax this assumption, and introduce semi-Markov and non-Markovian processes, respectively. The case where the probability distributions become stationary in the long-time limit is discussed in section 1.4, where we also discriminate between *equilibrium* and *non-equilibrium* situations. In section 1.6 we introduce the large deviation approach to non-equilibrium stationary states. Finally, some relevant numerical methods are presented in section 1.7.

1.1 Discrete-time Markov processes

We define a discrete-time Markov process as a sequence of random variables

$$X(0), X(1), X(2), \ldots \tag{1.1}$$

with the property that the probability of having a certain realisation of $X(n)$, given the values for $X(0), X(1), \ldots, X(n-1)$, depends only on the value assumed by $X(n-1)$, but not on $X(0), X(1), \ldots, X(n-2)$, i.e.,

$$\begin{aligned}\text{Prob}\{X(n) = x_n | X(n-1) = x_{n-1}, \ldots, X(1) = x_1, X(0) = x_0\} \\ = \text{Prob}\{X(n) = x_n | X(n-1) = x_{n-1}\}.\end{aligned} \tag{1.2}$$

This means that, for the conditional distribution of $X(n)$, only the value assumed by $X(n-1)$ is relevant. Roughly speaking, we can think that, after each step, the events happened in the past do not contribute to the subsequent evolution, as if they were forgotten. When the property (1.2) is not valid the stochastic process is said to be non-Markovian. It is worth saying that the outcomes $x_1, x_2, \ldots$ are elements of $\mathcal{S}$, but we can safely think of them just as natural numbers, as we are dealing with discrete $\mathcal{S}$. Such a notation with subscripts has been chosen as we will need to consider ordered sequences of configurations $(x_1, x_2, x_3, \ldots, x_n)$ corresponding to realisations of the process[1].

The sequence (1.1) with the property (1.2) can describe the following physical situation. At a given integer-valued time t, the configuration adopted by a system is assumed

[1] We remark that, in a slight shift in notation, we will also use x_i as a generic configuration label independently of time.

to be x_i. At time $t + \Delta t$ an instantaneous transition to a configuration x_n occurs with time-independent probability T_{x_{n+1},x_n}. The probability that no transition occurs, leaving the configuration unchanged, is $T_{x_{n+1},x_n} = 1 - \sum_{x_{n+1} \neq x_n} T_{x_{n+1},x_n}$. Applying this rule to an initial state x_0 generates the *trajectory*. Without loss of generality, the interval Δt is assumed to be equal to one.

In finite configuration space, it is convenient to group the probabilities T_{x_j,x_i} of jumps from x_i to x_j (x_j and x_j being labels for generic configurations) into a matrix $\mathbf{T}$ with entries $[\mathbf{T}]_{x_j,x_i} = T_{x_j,x_i}$. By definition, such a matrix has only non-negative entries, i.e. it is *non-negative* and is represented by a weighted *directed graph* or *digraph* $\mathcal{G}(\mathbf{T})$. This is the set of the vertices corresponding to each element of $\mathcal{S}$ and the set of the edges directed from x_i to x_j with weight T_{x_j,x_i}, whenever such a weight is non-zero. Within this framework, a trajectory of a stochastic process is equivalent to a *walk*, i.e., a sequence of edges that connect a sequence of vertices.

A digraph is called *strongly connected* if for each pair of vertices x_j, x_i, a walk from x_j to x_i and a walk from x_j to x_i both exist (if there are walks only in one of the directions, then the digraph is simply said to be *connected*). The important notion of *irreducibility* of $\mathbf{T}$ is closely related to that the connectedness of $\mathcal{G}(\mathbf{T})$. Specifically, a matrix $\mathbf{T}$ is said to be irreducible[2] if $\mathcal{S}$ cannot be divided into two disjoint non-empty sets $\mathcal{S}_1$ and $\mathcal{S}_2$ such that $[\mathbf{T}]_{x_i,x_j} = 0$ for all $x_i \in \mathcal{S}_1$ and $x_j \in \mathcal{S}_2$. In addition to this, $\mathbf{T}$ is irreducible if and only if $\mathcal{G}(\mathbf{T})$ is strongly connected [52]. In an irreducible matrix, for each $x_i, x_j \in \mathcal{S}$, there exist an integer n such that $[\mathbf{T}^n]_{x_i,x_j} > 0$.

A stronger property is that of *primitivity*. A matrix is referred to as primitive if there exists at least one integer n such that $[\mathbf{T}^n]_{x_i,x_j} > 0$ *for all* configurations x_i and x_j simultaneously, and is called *imprimitive* otherwise. An important necessary and sufficient condition for $\mathbf{T}$ to be primitive is that it has only one real eigenvalue on the spectral radius[3]. The equivalent graphical condition is that there is a walk of length n between every pair of configurations in $\mathcal{S}$.

Any outcome of the Markov chain (1.1) needs the initial probability mass distribution $\text{Prob}\{X_0 = x\} =: P_x(0)$ to be fixed. We define the column vectors

$$|P(0)\rangle = \sum_{x_i} P_{x_i}(0)|e_{x_i}\rangle \quad \text{and} \quad |P(n)\rangle = \sum_{x_i} P_{x_i}(n)|e_{x_i}\rangle, \tag{1.3}$$

$n = 1, 2, \ldots$, where $|e_{x_i}\rangle$ is the column vector with the x_i-th component equal to one

[2] An irreducible and non-negative matrix is also said to be *essentially positive*.

[3] When there are a number h of eigenvalues on the spectral radius, then the Markov chain is periodic with period h.

and the remaining components equal to zero, while $\text{Prob}\{X_n = x\} =: P_x(n)$. Hence, the probability mass at time n is given by $|P(n)\rangle = \mathbf{T}^n|P(0)\rangle$ while the probability distribution at time $n+1$ is related to that at time n through the Master equation $|P(n+1)\rangle = \mathbf{T}|P(n)\rangle$.

A special probability mass distribution is the so called *limiting distribution*, i.e.,

$$\lim_{n\to\infty} |P(n)\rangle. \tag{1.4}$$

An equally important distribution is the *invariant* distribution $|P^*\rangle$, i.e., a distribution that is not modified by the action of $\mathbf{T}$:

$$|P^*\rangle = \mathbf{T}|P^*\rangle. \tag{1.5}$$

We generically refer to both $\lim_{n\to\infty} |P(n)\rangle$ and $|P^*\rangle$ as a *stationary* solution or distribution and use the same symbol $|P^*\rangle$; in fact, for an irreducible primitive transition matrix the limiting distribution exists and is unique and is equal to the invariant distribution[4]. If an irreducible matrix is not primitive, then the limiting distribution does not exist. However, it is possible to see that the invariant distribution has the same interpretation in both primitive and imprimitive cases, i.e., the component of $|P^*\rangle$ is the long run fraction of the time that the chain spends in the corresponding configuration, see e.g., reference [52]. This, in turn, allows us to interchange time average and ensemble average, a property that in Physics is known as *ergodicity*[5].

Hereafter, when dealing with Markov processes, we will only consider irreducible primitive matrices (and strongly connected graphical representations). Moreover, in the next section we will consider continuous-time Markov chains, where periodicity is not permitted and it is not necessary to distinguish limiting from invariant distributions.

[4]This can be easily proven using the spectral representation and using the fact that the largest eigenvalue dominates the long-time limit, see, e.g., reference [34]. In this thesis we are mainly concerned with continuous-time processes and such a result is discussed in the next section.

[5]Hence irreducible chains are also said to be *ergodic*, although, in mathematical literature, this refers to a stronger property.

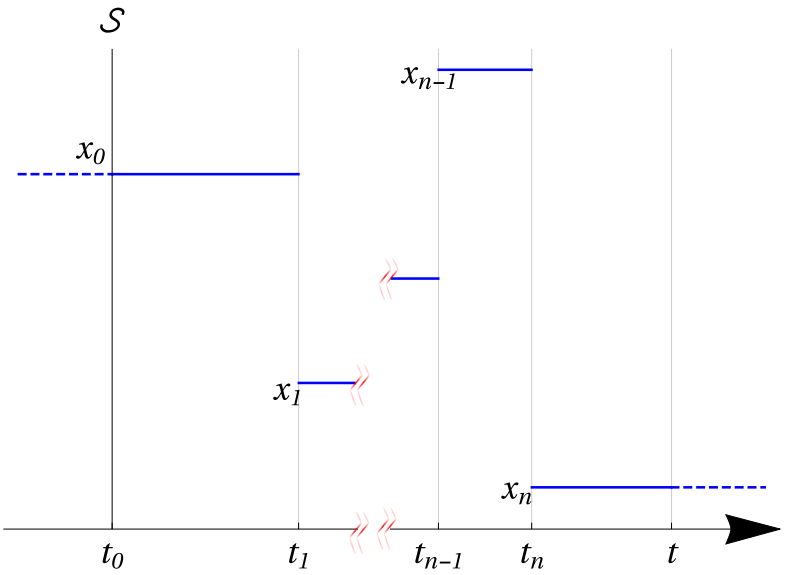

Figure 1.1: Representation of the trajectory $w(t)$ of equation (1.7).

1.2 Continuous-time Markov processes

We now consider a family $\{X(t) : t \geq 0\}$ of RVs in $\mathcal{S}$ indexed by a time $t \in [0, \infty)$. Such X is referred to as a continuous-time Markov process if

$$\text{Prob}\{X(t) = x_n | X(t_0) = x_0, \ldots, X(t_{n-1}) = x_{n-1}\} = \text{Prob}\{X(t) = x_n | X(t_{n-1}) = x_{n-1}\} \tag{1.6}$$

for all $x_0, \ldots x_{n-1}, x_n \in \mathcal{S}$ and any sequence $t_0 \leq \ldots \leq t_{n-1} \leq t_n \leq t$ of times. A convenient way to represent a portion of a trajectory $w(t)$ of duration $(t - t_0)$ starting in x_0 at t_0 is to write explicitly the sequence of instants where a jump occurs along with the arrival configurations (see figure 1.1):

$$w(t) := (t_0, x_0,\ t_1, x_1,\ \ \ldots\ t_{n-1}, x_{n-1},\ t_n, x_n,\ t). \tag{1.7}$$

We now use x_j and x_i as generic configuration labels. It is possible to prove (see reference [34]) that a continuous-time Markov process is defined when we have a set of functions $T_{x_i,x_j}(t)$, encoding for $\text{Prob}\{X(t_n) = x_i | X(t_{n-1}) = x_j\}$, with $t = t_n - t_{n-1}$,

that satisfy

$$0 \leq T_{x_i,x_j}(t) \leq 1 \tag{1.8}$$

$$T_{x_i,x_j}(t+s) = \sum_{x_k} T_{x_i,x_k}(t) T_{x_k,x_j}(s) \tag{1.9}$$

$$\sum_{x_i} T_{x_i,x_j}(t) \leq 1. \tag{1.10}$$

Equation (1.10) deserves some explanation: if the equality is satisfied for all x_j and t, the process is said to be *honest*, otherwise, if there is a strict inequality for some x_j and t, the process is said to be *dishonest*. The contribution necessary to complete the sum to one corresponds to the probability that the system escapes to infinity (if $\mathcal{S}$ is not a finite set) at a finite time. This situation would make the formulation of a general theory of continuous-time Markov processes quite involved. However, in this thesis, we will not consider such a pathological case, focusing on situations where the system can only escape to infinity in the limit as the time approaches infinity. We now make the additional assumption that the functions $T_{x_i,x_j}(t)$ are continuous in time. It follows that, over short intervals $\mathrm{d}t$,

$$T_{x_i,x_j}(\mathrm{d}t) = g_{x_i,x_j}\mathrm{d}t + o(\mathrm{d}t) \quad (x_i \neq x_j) \tag{1.11}$$

$$T_{x_i,x_i}(\mathrm{d}t) = 1 - g_{x_i}\mathrm{d}t + o(\mathrm{d}t) \tag{1.12}$$

The quantities g_{x_i,x_j} and g_{x_i} are referred to as *transition rates* and are typically grouped into a matrix called *generator* of the process whose entries are $[\mathbf{G}]_{x_i,x_j} = g_{x_i,x_j}$ (if $x_i \neq x_j$) or $[\mathbf{G}]_{x_i,x_i} = -\sum_{x_j} g_{x_j,x_i} =: -g_{x_i}$; a used convention is to set $g_{x_i,x_i} = 0$. If $[\mathbf{G}]_{x_i,x_i} = 0$, the state x_i is called *absorbing*, while if $[\mathbf{G}]_{x_i,x_i} = -\infty$, it is called *instantaneous*. Similarly, the quantities $T_{x_i,x_j}(t)$ are grouped into a family of matrices $\{\mathbf{T}(t)\}$ that forms a *standard semigroup*. A semigroup $\{\mathbf{T}(t)\}$ is called standard if $\mathbf{T}(t) \to \mathbb{1}$ as $t \to 0$. In the limit as $\mathrm{d}t \to 0$, equations (1.11) and (1.12) can be written as the matrix form

$$\lim_{\mathrm{d}t\to 0} \frac{\mathbf{T}(\mathrm{d}t) - \mathbb{1}}{\mathrm{d}t} = \mathbf{G}. \tag{1.13}$$

Dealing with standard semi-groups rules out dishonest processes. We can also impose a stronger condition on $\{\mathbf{T}(t)\}$ and say that a semi-group is *uniform* if $\{\mathbf{T}(t)\} \to \mathbb{1}$ uniformly as $t \to 0$. A theorem guarantees that $\{\mathbf{T}(t)\}$ is uniform if and only $\sup_i [G]_{ii} < \infty$, which also rules out instantaneous states [34]. If we define the column vector $|P(t)\rangle$ analogously to the discrete-time case, from equation (1.13), along with $|P(t+\mathrm{d}t)\rangle = \mathbf{T}(\mathrm{d}t)|P(t)\rangle$, we get the *Master equation*

$$\frac{\mathrm{d}}{\mathrm{d}t}|P(t)\rangle = \mathbf{G}|P(t)\rangle, \qquad |P(t)\rangle = \exp(t\mathbf{G})\,|P(0)\rangle. \tag{1.14}$$

Equation (1.9) is the Chapman-Kolmogorov equation for a Markov process. Taking the limit as $s \to 0$ and using equations (1.11) and (1.12), we obtain the *Kolmogorov forward equations*

$$\frac{\mathrm{d}\mathbf{T}(t)}{\mathrm{d}t} = \mathbf{G}\,\mathbf{T}(t); \tag{1.15}$$

a similar argument, using the limit as $t \to 0$, gives the *Kolmogorov backward equations*

$$\frac{\mathrm{d}\mathbf{T}(t)}{\mathrm{d}t} = \mathbf{T}(t)\,\mathbf{G}. \tag{1.16}$$

It is possible to prove that when $\{\mathbf{T}(t)\}$ forms a uniform semigroup with generator $\mathbf{G}$, then

$$\mathbf{T}(t) = \exp(t\mathbf{G}) \tag{1.17}$$

is the unique solution of the equations (1.15) and (1.16), subjected to the initial condition $\mathbf{T}(0) = \mathbb{1}$ [34]. For non-uniform semigroups, we will assume that equation (1.17) remains valid with $\exp(t\mathbf{G}) = \sum_{n=0}^{\infty} \mathbf{G}^n t^n/n!$. As a consequence, the invariant distribution satisfies $\mathbf{G}|P^*\rangle = 0$.

We now sketch some spectral theory. First of all, we consider a process with finite number of states and generator $\mathbf{G}$. Then it is possible to express $\mathbf{G}$ as

$$\mathbf{G} = \mathbf{\Psi}\mathbf{\Lambda}\mathbf{\Psi}^{-1}, \tag{1.18}$$

where $\mathbf{\Lambda}$ is diagonal, the columns of $\mathbf{\Psi}$ are right eigenvectors of $\mathbf{G}$, while the rows of $\mathbf{\Psi}^{-1}$ are left eigenvectors of $\mathbf{G}$. The diagonal entries of $\mathbf{\Lambda}$ are the eigenvalues λ_i of $\mathbf{G}$. Clearly, the set of eigenvectors form a basis in $\mathcal{S}$, i.e.,

$$\mathbf{\Psi}\mathbf{\Psi}^{-1} = \mathbb{1}. \tag{1.19}$$

From equation (1.17) we have

$$\mathbf{T}(t) = \mathbf{\Psi}\exp(\mathbf{\Lambda}t)\mathbf{\Psi}^{-1}. \tag{1.20}$$

Now, all the off-diagonal entries of $\mathbf{G}$ are non-negative, then, there exists a scalar k such that $\mathbf{M} = \mathbf{G} + k\mathbb{1}$ has only non-negative entries; consequently $\mathrm{e}^{t\mathbf{G}} = \mathrm{e}^{-kt}\mathrm{e}^{t\mathbf{M}}$. The first factor is positive and we only need to study the second factor $\mathrm{e}^{t\mathbf{M}} = \sum_{n=0}^{\infty} \mathbf{M}^n t^n/n!$. Now if we assume that $\mathbf{M}$ is irreducible (see section 1.1) we can always find an integer n such that $[\mathbf{M}^n]_{x_i,x_j} > 0$ for each pair (x_i, x_j). Since all positive integer powers of $\mathbf{M}$ appear in such series expansion, it follows that $\mathrm{e}^{t\mathbf{M}}$ is strictly positive, and so is $\mathrm{e}^{t\mathbf{G}}$, thus having only one real eigenvalue on the spectral radius (this is part of the Perron–Frobenius theory of non-negative matrices, see, e.g., reference [52]). This excludes

periodic behaviour in continuous time and ensures ergodicity (in the sense of section 1.1). The largest eigenvalue governs the long-time behaviour of $\mathrm{e}^{t\mathbf{G}}$, as shown in the following. Explicitly, the matrices $\mathbf{\Psi}$ and $\mathbf{\Psi}^{-1}$ can be expressed as

$$\mathbf{\Psi} = \sum_i |\psi_i\rangle\langle e_i|, \tag{1.21}$$

$$\mathbf{\Psi}^{-1} = \sum_i |e_i\rangle\langle\psi_i|, \tag{1.22}$$

where $\langle e_i|$ is the row vector with zero entries everywhere except for the i-th entry, which is one, while $|\psi_i\rangle$ and $\langle\psi_i|$ are the i-th eigenvectors, right and left respectively, of $\mathbf{G}$. Using these, it is possible to write equation (1.20) as

$$\mathbf{T}(t) = \sum_{i,j} |\psi_i\rangle\langle e_i|\mathrm{e}^{t\lambda_i}|e_j\rangle\langle\psi_j| = \sum_i |\psi_i\rangle\langle\psi_i|\mathrm{e}^{t\lambda_i}, \tag{1.23}$$

where we made use of the Kronecker delta $\langle e_i|e_j\rangle = \delta_{i,j}$. It appears clear that the dominant term, as t approaches infinity, is that with the largest eigenvalue, which we call λ_1. We now recall that the columns of $\mathbf{G}$ sum to zero, which implies that $\lambda_1 = 0$ and that the row vector $\langle 1|$ with all entries equal to one is the left eigenvector associated with λ_1 (i.e., $\langle 1|\mathbf{G} = 0$). As a consequence, we get $\lim_{t\to\infty} T(t)|P(0)\rangle = |\psi_1\rangle$, i.e., the leading right eigenvector is the limiting distribution for the Markov process. Clearly, being an eigenvector, this is also invariant.

Developing a general spectral theory in the case with infinite configuration space is more difficult as the eventuality of a continuous spectrum must be taken into account. We only mention here that an elegant spectral representation has been derived by S. Karlin and J. L. McGregor [47] for a special but ubiquitous model, viz., the *generalised birth-death process*. Such a model can be defined by its generator, i.e.,

$$\mathbf{G} = \begin{pmatrix} -\alpha_0 & g_1 & 0 & 0 & \dots \\ \alpha_0 & -\alpha_1 - g_1 & g_2 & 0 & \\ 0 & \alpha_1 & -\alpha_2 - g_2 & g_3 & \\ \vdots & & & & \ddots \end{pmatrix}. \tag{1.24}$$

Its spectral representation is written in terms of the polynomials $Q_x(\lambda)$, which are solutions of

$$T_{ij}(t) = w_j \int_0^\infty \mathrm{e}^{t\lambda} Q_i(\lambda) Q_j(\lambda)\, \mathrm{d}\mu(\lambda). \tag{1.25}$$

For such a system of equations, λ is real and $Q_i(\lambda)$ satisfies

$$w_j \int_0^\infty Q_i(\lambda) Q_j(\lambda)\, \mathrm{d}\mu(\lambda) = \delta_{ij}, \tag{1.26}$$

$$Q_0(\lambda) = 1, \qquad g_n Q_{n-1} - (\lambda + \alpha_n + g_n) Q_n(\lambda) + \alpha_n Q_{n+1}(\lambda) = 0, \tag{1.27}$$

$$w_0 = 1, \qquad w_j = \frac{\alpha_0, \alpha_1, \ldots, \alpha_{j-1}}{g_1, g_2, \ldots, g_j}. \tag{1.28}$$

The main difficulty in applying such a result is in deriving $\mu(\lambda)$. In appendix A the spectrum of a tridiagonal operator with constant entries is derived in detail, even without imposing $\langle 1|\mathbf{G} = 0$.

It is worth mentioning that, in the simple case where all the upper diagonal terms of the generator (1.24) are null while its lower diagonal entries are constant α, the birth-death process reduces to a "pure birth" process. For this case, one is only interested in the sequence of inter-event times—the sequence of configurations being trivial. This defines a new process, called the *Poisson process* of *intensity* α, which is one of the simplest and best known *renewal processes* [11].

All the continuous-time processes considered so far involve inter-event times that are exponentially distributed. To see this, we study the joint probability $\psi_{x_j,x_i}(t)\, \mathrm{d}t$ defined as the probability that the transition $x_i \to x_j$ occurs during the infinitesimal interval $[t, t+\mathrm{d}t)$ and the configuration x_i survives until t. In terms of random variables, $\psi_{x_j,x_i}(t)$ is the joint density function for the two random variables "time to the next jump" *and* "next configuration". We refer to it as the *waiting-time density* distribution (WTD). As we decided to focus on discrete configurations $x_i \in \mathcal{S}$ throughout this thesis, $\psi_{x_j,x_i}(t)$ obeys the following normalization condition

$$\sum_{x_j} \int_0^\infty \psi_{x_j,x_i}(t) \mathrm{d}t = 1, \quad \forall x_i. \tag{1.29}$$

We derive an explicit formula for $\psi_{x_j,x_i}(t)$ under the assumptions (1.11) and (1.12). If we imagine that the time interval $[0, t)$ is subdivided into k subintervals of length t/k, where $k > 1$, for the probability that the move to x_j occurs at the end of such an interval, while nothing happens before, can be written as:

$$\psi_{x_j,x_i}(t)\, \mathrm{d}t = [1 - g_{x_i} t/k + o(t/k)]^k\, [g_{x_j,x_i} \mathrm{d}t + o(\mathrm{d}t)]. \tag{1.30}$$

Then, we divide by $\mathrm{d}t$, and take the limit $\mathrm{d}t \to 0$, we obtain

$$\psi_{x_j,x_i}(t) = [1 - g_{x_i} t/k + o(t/k)]^k\, g_{x_j,x_i}. \tag{1.31}$$

In the limit as $k \to \infty$, as the first factor is simply $\exp(-g_{x_i}\, t)$, we obtain

$$\psi_{x_j,x_i}(t) = g_{x_j,x_i} \exp(-g_{x_i} t). \tag{1.32}$$

Marginalising over all arrival configurations, we obtain that the time spent in a certain state follows the exponential distribution $\psi_{x_i}(t) = g_{x_i} \exp(-g_{x_i}\, t)$, which indeed has the memory-less property. In discrete time, a similar argument shows that the duration of stay in a state has a geometric distribution.

When the transition rates are functions of time we obtain a *non-homogeneous* Markov process and we can no longer write the transition probabilities only as function of only t, but instead

$$T_{x_j,x_i}(t,s) = \text{Prob}\{X(s+t) = x_j | X(s) = x_i\}. \tag{1.33}$$

The Kolmogorov forward and backwards equations can be generalised to this case, i.e.,

$$\frac{\partial \mathbf{T}(t,s)}{\partial t} = \mathbf{G}(t)\, \mathbf{T}(t,s), \tag{1.34}$$

$$\frac{\partial \mathbf{T}(t,s)}{\partial s} = -\mathbf{T}(t,s)\, \mathbf{G}(s), \tag{1.35}$$

while dealing with their general formal solutions is often not straightforward as the inter-event times are, in general, non-exponentially distributed [12]. However such processes can be regarded as Markov as the time is obviously known at any instant and it is thus possible to make predictions on future states based on the present.

In the next section we tackle a different class of stochastic processes whose time evolution can be written in terms of rates that depend on a random time (contrarily to the time involved in non-homogeneous Markov processes).

1.3 Semi-Markov processes

In this section, we deal with a general class of continuous-time stochastic processes, where the jumps occur after times which are not necessarily exponentially distributed, and the probability of having a certain transition can depend on the time elapsed since the last jump. This means that memory is lost after each jump, but not during the jump. In mathematical literature such processes are referred to as *semi-Markov*, while in physics they are called *continuous-time random walks* (CTRWs). These were introduced in physics to model transport on lattices [37, 53], and later used in many other contexts, e.g., to describe quantum dots [19], temporal networks [39], animal movements [31], biochemical reactions [21, 22], and single-molecule kinetics [55, 67], and financial markets [50].

1.3.1 General waiting-time distribution

To develop a theory of stochastic processes when the distribution of life-times are not exponential, three main methods are available:

1) A first rough device is to the define an *embedded* Markov chain. This consists of a new Markov chain on a coarse-grained set of time instants, equally spaced. The interval between two consecutive instants is chosen so that the memory of the previous instant becomes irrelevant. This method is not rigorous and carries all the advantages/disadvantages of dealing with discrete-time Markov processes. This thesis is not focused on embedded Markov chains (although some general results of chapter 4 naturally apply to such cases).

2) A second antithetic solution consists of dividing the inter-event times into fictitious stages, such that the time spent in each stage has an exponential distribution. The resulting process is Markovian, but on an extended configuration space: in order to define the process, we need to specify both the configuration and the stage of life it has been reached. This device will be used to exploit some established results from the theory of Markov processes.

   An important case arises when a number k of stages are arranged in series. The total waiting time for this case is the sum of k exponentially distributed waiting times and has density distribution $\psi(\tau)$ neatly expressed through its Laplace transform

$$\int_0^\infty \mathrm{e}^{-\nu\tau}\psi(\tau)\,\mathrm{d}\tau = \prod_{i=1}^{k} \frac{g_i}{g_i+\nu}, \tag{1.36}$$

   where g_i is the rate associated with the stage i. A WTD satisfying equation (1.36) is referred to as an Hypoexponential distribution, which represents an Erlang distribution when all the rates g_i are equal [61].

   More generically, it has been proved in [10] that any probability density distribution having a rational Laplace transform $\overline{f}(\nu)$, with k poles and numerator of degree at most k, can be reproduced by a sequence of k exponential phases, not necessarily arranged in series. The probability distributions with this property are called *Coxian* (a subclass of *phase-type* distributions). Without loss of generality, we can make the following partial fraction decomposition

$$\overline{f}(\nu) = p_0 + q_0 p_1 \frac{\lambda_1}{\nu+\lambda_1} + \sum_{i=2}^{k} q_0 \dots q_{i-1} p_i \prod_{l=1}^{i} \frac{\lambda_l}{\nu+\lambda_l}, \tag{1.37}$$

where the poles are at $-\lambda_i$, $i = 1, 2, \ldots, k$, $p_{i-1} + q_{i-1} = 1$ and $p_k = 1$. Equation (1.37) has a simple interpretation in the time domain. At each stage $i-1$, there is a probability p_{i-1} of immediate escape and a probability q_{i-1} of entering the stage i, whose WTD is exponential with rate λ_i. A thorough account of such distributions can be found e.g., in W. J. Stewart [61].

3) A third option is to include a *continuous* supplementary variable called *age* or *expended life-time*, which is measured from the instant when the last event occurred. This also reflects the cases where a system is Markovian, but we do not have access to the microscopic mechanisms involved in the dynamics of the system, but we have instead information about the statistics of the times between jumps. This case is more general than the devices of 1) and 2). D. R. Cox [10] argued that WTDs with continuous supplementary variable have rational Laplace transforms with complex poles.

We now focus on WTDs of the case 3). For convenience, we set $\psi_{x_i,x_i}(\tau) = 0$. The WTDs $\psi_{x_j,x_i}(\tau)$, where $x_j, x_i \in \mathcal{S}$, fully describe the dynamics of a semi-Markov stochastic process in continuous time on a discrete set of configurations $\mathcal{S}$, provided that the initial state has been reached exactly at time t_0. Otherwise, we need either to know the time elapsed since the last jump or consider that the first WTD can be different from the others. The probability density of observing a trajectory (1.7) hence is

$$\varrho[w(t)] = \phi_{x_n}(t - t_n)\psi_{x_n,x_{n-1}}(t_n - t_{n-1}) \ldots \psi'_{x_1,x_0}(t_1 - t_0)P_{x_0}(t_0). \tag{1.38}$$

A natural situation is when we observe a portion of a trajectory that started before t_0. In this case, $\psi'_{x_1,x_0}(t_1 - t_0) = \psi_{x_1,x_0}(t_1 - t_{-1})/\phi_{x_0}(t_0 - t_{-1})$, as it depends on the time t_{-1} of the last jump before t_0, being conditioned on the survival until t_0. Also, in this case, $\phi'_{x_i}(t) = \phi_{x_i}(-t_{-1} + t)/\phi_{x_i}(-t_{-1})$ where the probability that x_i survives the interval $t - t_0$ is $\phi'_{x_i}(t - t_0) = \phi_{x_i}(t - t_{-1})/\phi_{x_i}(t_0 - t_{-1})$.

The WTDs $\psi_{x_i,x_{i-1}}(\tau)$ are joint probability densities for the transition age and the destination state, represented by the variables τ and x_i, respectively:

$$\psi_{x_i,x_{i-1}}(\tau) = \text{Prob}\{\text{the jump is } x_{i-1} \to x_i \text{ and the transition is at age } \tau\}. \tag{1.39}$$

A useful convention is to write the WTD in the form

$$\psi_{x_j,x_i}(\tau) = p_{x_j,x_i}(\tau) \times \psi_{x_i}(\tau), \tag{1.40}$$

with the normalization conditions $\sum_{x_j} p_{x_j,x_i}(\tau) = 1$ and $\int_0^\infty \psi_{x_i}(\tau)\mathrm{d}\tau = 1$. Explicitly equation (1.40) reads:

$$\begin{aligned}&\text{Prob}\{\text{the jump is } x_i \to x_j \text{ and the transition is at age } \tau\}\\ &= \text{Prob}\{\text{the jump is } x_i \to x_j | \text{the transition is at age } \tau\} \times \text{Prob}\{\text{the transition is at age } \tau\}.\end{aligned} \tag{1.41}$$

We will also find useful the following definition,

$$\mathsf{P}_{x_j,x_i} = \int_0^\infty \psi_{x_j,x_i}(\tau)\mathrm{d}\tau, \tag{1.42}$$

which is the probability that, given that the system is in state x_i, the next jump will be to state x_j. This gives us an alternative factorisation for the WTD,

$$\psi_{x_j,x_i}(t) = \mathsf{P}_{x_j,x_i} \times f_{x_j,x_i}(\tau), \tag{1.43}$$

where $f_{x_j,x_i}(t)$ is the conditional probability that the transition is at age τ, given that the configuration changes from x_i to x_j—this can also be expressed with $f_{x_j,x_i}(t) = \int_0^t \psi_{x_j,x_i}(\tau)\mathrm{d}\tau / p_{x_j,x_i}$. Hence, equation (1.43) reads:

$$\begin{aligned}&\text{Prob}\{\text{the jump is } x_i \to x_j \text{ and the transition is at age } \tau\}\\ &= \text{Prob}\{\text{the jump is } x_i \to x_j\} \times \text{Prob}\{\text{the transition is at age } \tau | \text{the jump is } x_i \to x_j\}.\end{aligned} \tag{1.44}$$

If the two events of equation (1.39) are independent, we can obviously write

$$\begin{aligned}&\text{Prob}\{\text{the jump is } x_i \to x_j \text{ and the transition is at age } \tau\} =\\ &\qquad \text{Prob}\{\text{the jump is } x_i \to x_j\} \times \text{Prob}\{\text{the transition is at age } \tau\}\end{aligned} \tag{1.45}$$

and say that the semi-Markov process enjoys direction–time independence (DTI) [3, 7, 55, 67]. This implies that it is possible to eliminate the dependence on the time in the jump-dependent factor of the PDF (1.40), i.e.,

$$\psi_{x_j,x_i}(\tau) = p_{x_j,x_i}\psi_{x_i}(\tau), \tag{1.46}$$

and eliminate the dependence on the next configuration x_j in $f_{x_j,x_i}(\tau)$, i.e.,

$$\psi_{x_j,x_i}(\tau) = \mathsf{P}_{x_j,x_i} f_{x_i}(\tau) \tag{1.47}$$

Hence, in presence of DTI, p_{x_j,x_i} and $\psi_{x_i}(\tau)$ coincide with P_{x_j,x_i} and $f_{x_i}(\tau)$, respectively. More interesting consequences of the DTI condition are discussed in section 1.4.2.

The important quantity

$$\psi_{x_i}(\tau) = \sum_{x_j} \psi_{x_j,x_i}(\tau) \tag{1.48}$$

is referred to as the *residence time distribution* (RTD) of the configuration x_i. Other ubiquitous quantities are the *survival probability*[6] and the cumulative distribution, defined respectively as

$$\phi_{x_i}(t) = \int_t^\infty \psi_{x_i}(\tau)d\tau \tag{1.49}$$

and

$$\phi^c_{x_i}(t) = \int_0^t \psi_{x_i}(\tau)d\tau. \tag{1.50}$$

Defining such probabilities is clearly redundant as they carry the same amount of information as $\psi_{x_i}(\tau)$ and are related by the following equations:

$$\psi_{x_i}(\tau) = \frac{\mathrm{d}\phi^c_{x_i}(\tau)}{\mathrm{d}\tau}, \tag{1.51}$$

$$\psi_{x_i}(\tau) = -\frac{\mathrm{d}\phi_{x_i}(\tau)}{\mathrm{d}\tau}. \tag{1.52}$$

Equation (1.49) is often encountered in actuarial mathematics, reliability engineering, and survival analysis [8], where one is interested in the probability that a state survives for a guarantee period.

The WTD $\psi_{x_i,x_j}(\tau)$ can also be expressed in terms of the age-specific *rate* or *hazard function* $g_{x_i,x_j}(\tau)$, which is the probability density that there is a transition $x_j \to x_i$ during the "infinitesimal" interval $[\tau, \tau+\mathrm{d}t)$, conditioned on having no transitions during the interval $[0, \tau)$:

$$\psi_{x_i,x_j}(\tau) = g_{x_i,x_j}(\tau)\phi_{x_j}(\tau), \tag{1.53}$$

where τ is measured from the instant when the last jump occurred. The meaning of the hazard function appears particularly clear if we consider the survival probability of the state x_j and set $g_{x_i,x_i}(\tau) = 0$. Let us sum (1.53) over $x_i \in \mathcal{S}$,

$$\psi_{x_j}(\tau) = \sum_{x_i} g_{x_i,x_j}(\tau)\phi_{x_j}(\tau), \tag{1.54}$$

[6] also known as *survivor function*.

and refer to the sum $\sum_{x_i} g_{x_i,x_j}(\tau)$ as the *age-specific* escape rate of x_j. Because of equation (1.52), we can write equation (1.54) in the form of a logarithmic derivative

$$\sum_{x_i} g_{x_i,x_j}(\tau) = -\frac{\mathrm{d} \ln \phi_{x_j}(\tau)}{\mathrm{d}\tau}, \tag{1.55}$$

and, integrating with initial condition $\phi_{x_i}(0) = 1$, we get

$$\phi_{x_j}(\tau) = \exp\left(- \int_0^\tau \sum_{x_i} g_{x_i,x_j}(t)\, \mathrm{d}t \right), \tag{1.56}$$

$$\psi_{x_j}(\tau) = \sum_{x_i} g_{x_i,x_j}(\tau) \exp\left(- \int_0^\tau \sum_{x_i} g_{x_i,x_j}(t)\, \mathrm{d}t \right), \tag{1.57}$$

which allows us to cast equation (1.53) in the form of equation (1.40)

$$\psi_{x_i,x_j}(\tau) = \frac{g_{x_i,x_j}(\tau)}{\sum_{x_i} g_{x_i,x_j}(\tau)} \psi_{x_j}(\tau), \tag{1.58}$$

which will be shown to be more convenient.

1.3.2 The exponential survival function

The special choice (1.32) gives the following RTD, cumulative distribution, and survival probability:

$$\psi_{x_j}(\tau) = \left(\sum_{x_i} g_{x_i,x_j} \right) \times \mathrm{e}^{-\sum_{x_i} g_{x_i,x_j} \tau}, \tag{1.59}$$

$$\phi^c_{x_j}(\tau) = 1 - \mathrm{e}^{-\sum_{x_i} g_{x_i,x_j} \tau}, \tag{1.60}$$

$$\phi_{x_j}(\tau) = \mathrm{e}^{-\sum_{x_i} g_{x_i,x_j} \tau}. \tag{1.61}$$

According to equations (1.56) and (1.57), these relations imply constant hazards $g_{x_i,x_j}(t) = g_{x_i,x_j}$, which have been simply referred to as *rates* in section 1.2. Equation (1.59) reminds us that the interval between events are *always* exponential in homogeneous Markov processes, while equation (1.61) makes evident another reason why memory-less continuous-time processes are typically simpler than non-Markov processes, i.e., their survival probabilities always factorise.

1.3.3 Generalised Master equation

We now introduce a device known as *generalised Master equation* (GME), which has been largely used in physics literature to deal with semi-Markov processes. The Master equation (1.14) is a convenient description of a process that entails the memory-less property; although it can describe complex trajectories in the configuration space $\mathcal{S}$, its dependence on time is simple, as prescribed by equations (1.11) and (1.12). The generalised Master equation is similar to the standard Master equation, but can be employed to describe the time evolution of the configuration probability vector $|P(t)\rangle$ in systems with non-exponentially distributed inter-events times.

We set the general formalism along the lines of references [19, 37, 53]. For a semi-Markov process, represented by equation (1.38), the probability of having a configuration x_i after a time $t - t_0$ since the reference instant t_0 can be explicitly written as the sum

$$
\begin{aligned}
P_{x_i}(t) &= \sum_{x_0} \delta_{x_i,x_0} \phi'_{x_0}(t-t_0) P_{x_0}(t_0) \qquad (1.62)\\
&+ \int_{t_0}^{t} \mathrm{d}t_1 \sum_{x_0,x_1} \delta_{x_i,x_1} \phi_{x_1}(t-t_1) \psi'_{x_1,x_0}(t_1-t_0) P_{x_0}(t_0)\\
&+ \int_{t_0}^{t} \mathrm{d}t_1 \int_{t_1}^{t} \mathrm{d}t_2 \sum_{x_0,x_1,x_2} \delta_{x_i,x_2} \phi_{x_2}(t-t_2) \psi_{x_2,x_1}(t_i-t_1) \psi'_{x_1,x_0}(t_1-t_0) P_{x_0}(t_0)\\
&+ \int_{t_0}^{t} \mathrm{d}t_1 \int_{t_1}^{t} \mathrm{d}t_2 \int_{t_2}^{t} \mathrm{d}t_3 \sum_{x_0,\ldots,x_3} \delta_{x_i,x_3} \phi_{x_3}(t-t_3) \psi_{x_3,x_2}(t_3-t_2) \psi_{x_2,x_1}(t_2-t_1) \psi'_{x_1,x_0}(t_1-t_0) P_{x_0}(t_0) + \ldots ,
\end{aligned}
$$

where jumps to the same configuration as the departure one are excluded, i.e., $x_1 \neq x_0$, $x_2 \neq x_1, \ldots$. We now derive a standard recursive relation from equation (1.62), assuming the observation began at the initial time $t_0 \geq 0$. This use of such an initial time makes the case we consider slightly more explicit than the analogues seen in most of the literature (e.g., in references [19, 53]). By reversing the order of integration in the integrals of equation (1.62) we have

$$
\begin{aligned}
P_{x_i}(t) &= \phi'_{x_i}(t-t_0) P_{x_i}(t_0) \qquad (1.63)\\
&+ \int_{t_0}^{t} \mathrm{d}t_1 \, \phi_{x_i}(t-t_1) \sum_{x_0} \psi'_{x_i,x_0}(t_1-t_0) P_{x_0}(t_0)\\
&+ \int_{t_0}^{t} \mathrm{d}t_2 \, \phi_{x_i}(t-t_2) \int_{t_0}^{t_2} \mathrm{d}t_1 \sum_{x_1} \psi_{x_i,x_1}(t_2-t_1) \sum_{x_0} \psi'_{x_1,x_0}(t_1-t_0) P_{x_0}(t_0)\\
&+ \int_{t_0}^{t} \mathrm{d}t_3 \, \phi_{x_i}(t-t_3) \int_{t_0}^{t_3} \mathrm{d}t_2 \sum_{x_2} \psi_{x_i,x_2}(t_3-t_2) \int_{t_0}^{t_2} \mathrm{d}t_1 \sum_{x_1} \psi_{x_2,x_1}(t_2-t_1) \sum_{x_0} \psi'_{x_1,x_0}(t_1-t_0) P_{x_0}(t_0) + \ldots .
\end{aligned}
$$

We desire to group some of the terms in the r.h.s. of equation (1.63) under the integral operator $\int_{t_0}^{t} \mathrm{d}\tau \phi_{x_i}(t-\tau)\cdot$. We first define the probability density $\eta_{x_i,n}(t_{n+1}-t_0)$ that the system jumps onto the state x during the interval $[t_{n+1}, t_{n+1}+\mathrm{d}t)$ (i.e., immediately after a time $t_{n+1}-t_0$ from the beginning of the observation) after n transitions. Using such a definition, the equation (1.63) can be written as

$$P_{x_i}(t) = \phi'_{x_i}(t-t_0)P_{x_0}(t_0) + \int_{t_0}^{t} \mathrm{d}t_1\, \phi_{x_i}(t-t_1)\eta_{x_i,0}(t_1-t_0) + \int_{t_0}^{t} \mathrm{d}t_2\, \phi_{x_i}(t-t_2)\eta_{x_i,1}(t_2-t_0) + \int_{t_0}^{t} \mathrm{d}t_3\, \phi_{x_i}(t-t_3)\eta_{x_i,2}(t_3-t_0) + \dots\,, \quad (1.64)$$

where it is clear that we are allowed to replace $t_1, t_2, \dots$ with τ . After such a replacement, the terms $\eta_{x_i,n}(\tau-t_0)$ can be grouped into the sum $\sum_{n=0}^{\infty}\eta_{x_i,n}(\tau-t_0) =: \eta_{x_i}(\tau-t_0)$, which represents the probability density that the system jumps onto the state x_i after a time $\tau-t_0$ since the beginning of the observation interval, regardless of the number of previous jumps. This leaves us with the convenient recursive equations:

$$P_{x_i}(t) = \phi'_{x_i}(t-t_0)P_x(t_0) + \int_{t_0}^{t} \phi_{x_i}(t-\tau)\eta_{x_i}(\tau-t_0)\,\mathrm{d}\tau, \quad (1.65)$$

$$\eta_{x_i}(t-t_0) = \sum_{x_j} \psi'_{x_i,x_j}(t-t_0)P_{x_j}(t_0) + \sum_{x_j}\int_{t_0}^{t} \psi_{x_i,x_j}(t-\tau)\eta_{x_j}(\tau-t_0)\,\mathrm{d}\tau. \quad (1.66)$$

It is convenient to make two changes of variable, by defining $u=\tau-t_0$ and $T=t-t_0$, which yield

$$P_{x_i}(T+t_0) = \phi'_{x_i}(T)P_x(t_0) + \int_{0}^{T} \phi_{x_i}(T-u)\eta_{x_i}(u)\,\mathrm{d}u, \quad (1.67)$$

$$\eta_{x_i}(T) = \sum_{x_j} \psi'_{x_i,x_j}(T)P_{x_j}(t_0) + \sum_{x_j}\int_{0}^{T} \psi_{x_i,x_j}(T-u)\eta_{x_j}(u)\,\mathrm{d}u. \quad (1.68)$$

Equation (1.68) can be expressed even more compactly after a Laplace transform, i.e.,

$$\overline{\eta}_{x_i}(\nu) = \sum_{x_j} \overline{\psi'}_{x_i,x_j}(\nu)P_{x_j}(t_0) + \sum_{x_j} \overline{\psi}_{x_i,x_j}(\nu)\overline{\eta}_{x_j}(\nu), \quad (1.69)$$

where

$$\overline{f}(\nu) = \int_{0}^{\infty} \mathrm{e}^{-\nu T} f(T)\,\mathrm{d}T, \quad (1.70)$$

and the convolution theorem

$$\int_{0}^{\infty} \mathrm{e}^{-\nu T}\left[\int_{0}^{T} f(T-u)g(u)\,\mathrm{d}u\right]\mathrm{d}T = \int_{0}^{\infty} \mathrm{e}^{-\nu T} f(T)\,\mathrm{d}T \times \int_{0}^{\infty} \mathrm{e}^{-\nu T} g(T)\,\mathrm{d}T, \quad (1.71)$$

has been used. The Laplace transform of equation (1.67) needs slightly more care. Using the notation

$$\widehat{f}(\nu) = \int_0^\infty \mathrm{e}^{-\nu T} f(T + t_0)\,\mathrm{d}T, \tag{1.72}$$

along with the convolution theorem for the r.h.s., we get

$$\widehat{P}_{x_i}(\nu) = \overline{\phi'}_{x_i}(\nu) P_{x_i}(t_0) + \overline{\phi}_{x_i}(\nu)\overline{\eta}_{x_i}(\nu). \tag{1.73}$$

From equation (1.71) it is straightforward to derive the formula for the Laplace transform of an integral:

$$\int_0^\infty \mathrm{e}^{-\nu t}\left[\int_0^t f(u)\,\mathrm{d}u\right]\mathrm{d}t = \frac{\int_0^\infty \mathrm{e}^{-\nu t} f(t)\,\mathrm{d}t}{\nu}. \tag{1.74}$$

This in turns gives the following explicit forms for the Laplace transform of the survival probabilities [11],

$$\overline{\phi}_{x_i}(\nu) = \frac{1 - \overline{\psi}_{x_i}(\nu)}{\nu}, \qquad \overline{\phi'}_{x_i}(\nu) = \frac{1 - \overline{\psi'}_{x_i}(\nu)}{\nu}. \tag{1.75}$$

which yield, along with equation (1.73),

$$\nu\widehat{P}_{x_i}(\nu) - P_{x_i}(t_0) = -\sum_{x_j} \overline{\psi'}_{x_j,x_i}(\nu) P_{x_i}(t_0) + \overline{\eta}_{x_i}(\nu) - \sum_{x_j} \overline{\psi}_{x_j,x_i}(\nu)\overline{\eta}_{x_i}(\nu). \tag{1.76}$$

Then, using (1.69) to substitute for the second term of the r.h.s., we get

$$\begin{aligned}\nu\widehat{P}_{x_i}(\nu) - P_{x_i}(t_0) = &-\sum_{x_j} \overline{\psi'}_{x_j,x_i}(\nu) P_{x_i}(t_0) + \sum_{x_j} \overline{\psi'}_{x_i,x_j}(\nu) P_{x_j}(t_0) \\ &+ \sum_{x_j} \overline{\psi}_{x_i,x_j}(\nu)\overline{\eta}_{x_j}(\nu) - \sum_{x_j} \overline{\psi}_{x_j,x_i}(\nu)\overline{\eta}_{x_i}(\nu).\end{aligned} \tag{1.77}$$

Plugging $\overline{\eta}_{x_i}(\nu)$ from equation (1.73) into the third and fourth terms on the r.h.s. of equation (1.77), we get the equation

$$\nu\widehat{P}_{x_i}(\nu) - P_{x_i}(t_0) = \overline{I}_{x_i}(\nu) + \sum_{x_j} \frac{\overline{\psi}_{x_i,x_j}(\nu)}{\overline{\phi}_{x_j}(\nu)}\widehat{P}_{x_j}(\nu) - \sum_{x_j} \frac{\overline{\psi}_{x_j,x_i}(\nu)}{\overline{\phi}_{x_i}(\nu)}\widehat{P}_{x_i}(\nu), \tag{1.78}$$

where $\overline{I}_{x_i}(\nu)$ contains the terms that explicitly depend on the WTDs at time t_0, i.e.,

$$\overline{I}_{x_i}(\nu) = -\sum_{x_j} \overline{\psi'}_{x_j,x_i}(\nu) P_{x_i}(t_0) + \sum_{x_j} \overline{\psi'}_{x_i,x_j}(\nu) P_{x_j}(t_0) - \sum_{x_j} \overline{\psi}_{x_i,x_j}(\nu) \frac{\overline{\phi'}_{x_j}(\nu)}{\overline{\phi}_{x_j}(\nu)} P_{x_j}(t_0) + \sum_{x_j} \overline{\psi}_{x_j,x_i}(\nu) \frac{\overline{\phi'}_{x_i}(\nu)}{\overline{\phi}_{x_i}(\nu)} P_{x_i}(t_0). \quad (1.79)$$

Integrating $\widehat{P}_{x_i}(\nu)$ by parts, we get that the l.h.s. of equation (1.78) is the Laplace transform of the time derivative of $P_{x_i}(t)$, in fact:

$$\nu \widehat{P}_{x_i}(\nu) - P_{x_i}(t_0) = \nu \left[\frac{\mathrm{e}^{-\nu T} P_{x_i}(T+t_0)}{-\nu} \right]_0^\infty - \nu \int_0^\infty \frac{1}{-\nu} \mathrm{e}^{-\nu T} \, \mathrm{d}P_{x_i}(T+t_0) - P_{x_i}(t_0) \quad (1.80)$$

$$= \int_0^\infty \mathrm{e}^{-\nu T} \frac{\mathrm{d}P_{x_i}(T+t_0)}{\mathrm{d}T} \, \mathrm{d}T. \quad (1.81)$$

We also define the *memory kernel* $K_{x_i,x_j}(t)$ through an equation similar to equation (1.53), but in the Laplace domain, i.e.,

$$\overline{\psi}_{x_i,x_j}(\nu) = \overline{K}_{x_i,x_j}(\nu) \overline{\phi}_{x_j}(\nu). \quad (1.82)$$

Applying to both sides of equation (1.78) the inverse Laplace transform we readily get the generalised Master equation,

$$\frac{\mathrm{d}}{\mathrm{d}T} P_{x_i}(T+t_0) = I_{x_i}(T) + \sum_{x_j} \int_0^T \left[K_{x_i,x_j}(T-\tau) P_{x_j}(\tau+t_0) - K_{x_j,x_i}(T-\tau) P_{x_i}(\tau+t_0) \right] \mathrm{d}\tau, \quad (1.83)$$

which can be written as

$$\frac{\mathrm{d}}{\mathrm{d}t} P_{x_i}(t) = I_{x_i}(t-t_0) + \sum_{x_j} \int_{t_0}^t \left[K_{x_i,x_j}(t-v) P_{x_j}(v) - K_{x_j,x_i}(t-v) P_{x_i}(v) \right] \mathrm{d}v, \quad (1.84)$$

using $T = t - t_0$ and defining $v = \tau + t_0$. The Markovian Master equation (1.14) is recovered when $K_{x_j,x_i}(t) = g_{x_j,x_i}\delta(t)$, which implies $\overline{K}_{x_j,x_i}(\nu) = g_{x_j,x_i}$ and, consequently, $\overline{I}_{x_i}(\nu) = 0$. Equation (1.84) incorporates the memory of events occurred before t_0 into the term $I_{x_i}(T)$, while memory kernels couple time instants following t_0. In the next section, we will show that, in the limit as $t \to \infty$, the term $I_{x_i}(t-t_0)$ becomes irrelevant for the probability-vector component $P_{x_i}(t)$.

1.4 Stationary states

This section deals with the long-time behaviour of a semi-Markov process, whose probability distribution is assumed to approach a stationary state.

1.4.1 The corresponding Markov process

As first step, we prove that, for many natural choices of WTDs,

$$\lim_{t\to\infty} I_{x_i}(t) = 0. \tag{1.85}$$

Following references [19, 60], let us first consider WTDs that have only finite moments, so that the following Maclaurin series expansion converges:

$$\begin{aligned}\overline{\psi}_{x_j,x_i}(\nu) &= \int_0^\infty \mathrm{e}^{-\nu t}\psi_{x_j,x_i}(t)\,\mathrm{d}t \\ &= \int_0^\infty \psi_{x_j,x_i}(t)\mathrm{d}t - \nu\int_0^\infty t\psi_{x_j,x_i}(t)\mathrm{d}t + \frac{\nu^2}{2}\int_0^\infty t^2\psi_{x_j,x_i}(t)\mathrm{d}t + \dots \\ &= \mathsf{P}_{x_j,x_i} - \nu\mathsf{A}_{x_j,x_i} + O(\nu^2),\end{aligned} \tag{1.86}$$

where the P_{x_j,x_i} and A_{x_j,x_i} are, respectively, the zeroth and first moments of $\psi_{x_j,x_i}(t)$, in this case. Alternatively, we consider α-stable distributions, defined by their Laplace transform

$$\begin{aligned}\overline{\psi}_{x_j,x_i}(\nu) &= \mathsf{P}_{x_j,x_i}\exp(-\nu^\alpha \mathsf{B}_{x_j,x_i}/\mathsf{P}_{x_j,x_i}) \\ &= \mathsf{P}_{x_j,x_i} - \nu^\alpha\mathsf{B}_{x_j,x_i} + O(\nu^{2\alpha}),\end{aligned} \tag{1.87}$$

where P_{x_j,x_i} and B_{x_j,x_i} are implicitly defined after expanding $\exp(-\nu^\alpha\mathsf{B}_{x_j,x_i}/\mathsf{P}_{x_j,x_i})$ and $0<\alpha<1$. This corresponds to WTDs that, in the time domain, decay as $\sim t^{-\alpha-1}$ and have infinite mean waiting times. In both cases (1.86) and (1.87), the limit as $\nu\to 0$ of $\overline{\psi}_{x_i,x_j}(\nu)$ and $\overline{\psi'}_{x_i,x_j}(\nu)$ can be represented by the algebraic forms $\mathsf{P}_{x_i,x_j} - \mathsf{B}_{x_i,x_j}\nu^\alpha$ and $\mathsf{P}'_{x_i,x_j} - \mathsf{B}'_{x_i,x_j}\nu^\alpha$, respectively. Using the standard relation (1.75) and setting $\mathsf{B}_{x_j} = \sum_{x_i}\mathsf{B}_{x_i,x_j}$ and $\mathsf{B}'_{x_j} = \sum_{x_i}\mathsf{B}'_{x_i,x_j}$, we get

$$\begin{aligned}\lim_{\nu\to 0}\overline{I}_{x_i}(\nu) = \lim_{\nu\to 0}\sum_{x_j}\Bigg[&-\left(\mathsf{P}'_{x_j,x_i} - \mathsf{B}'_{x_j,x_i}\nu^\alpha\right)\mathsf{P}_{x_i}(t_0) + \left(\mathsf{P}'_{x_i,x_j} - \mathsf{B}'_{x_i,x_j}\nu^\alpha\right)\mathsf{P}_{x_j}(t_0) \\ &- \left(\mathsf{P}_{x_i,x_j} - \mathsf{B}_{x_i,x_j}\nu^\alpha\right)\frac{\mathsf{B}'_{x_j}}{\mathsf{B}_{x_j}}\mathsf{P}_{x_j}(t_0) + \left(\mathsf{P}_{x_j,x_i} - \mathsf{B}_{x_j,x_i}\nu^\alpha\right)\frac{\mathsf{B}'_{x_i}}{\mathsf{B}_{x_i}}\mathsf{P}_{x_i}(t_0)\Bigg],\end{aligned} \tag{1.88}$$

which is finite in ν and implies

$$\lim_{\nu\to 0} \nu \overline{I}_{x_i}(\nu) = 0. \tag{1.89}$$

We finally assume that the limit $\lim_{t\to\infty} I_{x_i}(t)$ exists and is finite; then, the *final-value* theorem [20] ensures that this is equal to $\lim_{\nu\to 0} \nu \overline{I}_{x_i}(\nu)$, and the result (1.89) in turns implies equation (1.85).

This suggests that, to investigate the infinite-time solution of the generalised Master equation, we can neglect the initial-condition terms $I_{x_i}(t)$. Let us also assume that, in such an infinite-time limit, the probability vector reaches a stationary state

$$|P^*\rangle = \sum_{x_i} P^*_{x_i} |e_{x_i}\rangle. \tag{1.90}$$

This can be used to get rid of the memory kernel. In fact equation (1.84) implies

$$\frac{\mathrm{d}}{\mathrm{d}t} P_{x_i}(t) = I_{x_i}(t-t_0) + \sum_{x_j} \int_0^{t-t_0} \left[K_{x_i,x_j}(u) P_{x_j}(t-u) - K_{x_j,x_i}(u) P_{x_i}(t-u) \right] \mathrm{d}u; \tag{1.91}$$

which, in the limit as $t \to \infty$, yields

$$\sum_{x_j} P^*_{x_j} \int_0^\infty K_{x_i,x_j}(u)\,\mathrm{d}u - P^*_{x_i} \sum_{x_j} \int_0^\infty K_{x_j,x_i}(u)\,\mathrm{d}u = 0. \tag{1.92}$$

The integral $\int_0^\infty K_{x_i,x_j}(t)\,\mathrm{d}t$ is the Laplace transform of $K_{x_i,x_j}(t)$, computed at $\nu = 0$; consequently, from equation (1.82),

$$\overline{K}_{x_i,x_j}(0) = \frac{\overline{\psi}_{x_i,x_j}(0)}{\overline{\phi}_{x_j}(0)}. \tag{1.93}$$

The functions $\overline{\psi}_{x_i,x_j}(0)$ and $\overline{\phi}_{x_i,x_j}(0)$ have a neat physical meaning, as $\overline{\psi}_{x_i,x_j}(0) = \int_0^\infty \psi_{x_i,x_j}(t)\,\mathrm{d}t$ is the transition probability P_{x_i,x_j} seen in equation (1.42), while $\overline{\phi}_{x_i}(0) = \int_0^\infty \phi_{x_i}(t)\,\mathrm{d}t =: \mathsf{A}_{x_i}$ is the first moment of $\psi_{x_i}(t)$, i.e. the mean waiting time. Hence, we can write equation (1.92) as

$$\sum_{x_j} P^*_{x_j} \frac{\mathsf{P}_{x_i,x_j}}{\mathsf{A}_{x_j}} - P^*_{x_i} \sum_{x_j} \frac{\mathsf{P}_{x_j,x_i}}{\mathsf{A}_{x_i}} = 0, \tag{1.94}$$

which is the stationarity condition of a Markov process with rates $\mathsf{P}_{x_i,x_j}/\mathsf{A}_{x_j}$ and no memory kernel. Consequently, we can study the stationary state of a semi-Markov process by means of an effective continuous-time process, usually referred to as the *corresponding*

Markov Process, which has generator:

$$\begin{pmatrix} -\sum_i \frac{\mathsf{P}_{i,1}}{\mathsf{A}_1} & \frac{\mathsf{P}_{1,2}}{\mathsf{A}_2} & \frac{\mathsf{P}_{1,3}}{\mathsf{A}_3} & \frac{\mathsf{P}_{1,4}}{\mathsf{A}_4} & \cdots \\ \frac{\mathsf{P}_{2,1}}{\mathsf{A}_1} & -\sum_i \frac{\mathsf{P}_{i,2}}{\mathsf{A}_2} & \frac{\mathsf{P}_{2,3}}{\mathsf{A}_3} & \frac{\mathsf{P}_{2,4}}{\mathsf{A}_4} & \cdots \\ \frac{\mathsf{P}_{3,1}}{\mathsf{A}_1} & \frac{\mathsf{P}_{3,2}}{\mathsf{A}_2} & -\sum_i \frac{\mathsf{P}_{i,3}}{\mathsf{A}_3} & \frac{\mathsf{P}_{3,4}}{\mathsf{A}_4} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{1.95}$$

It is also convenient to define the stochastic matrix $\overline{\mathbf{K}}(\nu)$ with the off-diagonal entry x_i, x_j equal to the Laplace transform of the memory kernel $K_{x_i,x_j}(t)$

$$\begin{aligned} \left[\overline{\mathbf{K}}(\nu)\right]_{x_i,x_j} &= \overline{K}_{x_i,x_j}(\nu), \\ \left[\overline{\mathbf{K}}(\nu)\right]_{x_i,x_i} &= -\sum_{x_j} \overline{K}_{x_j,x_i}(\nu), \end{aligned} \tag{1.96}$$

This allows us to write the stationarity condition (1.94) in the compact form

$$\overline{\mathbf{K}}(0)|P^*\rangle = 0, \tag{1.97}$$

where $\overline{\mathbf{K}}(0)$ is equivalent to the generator (1.95). To lighten the notation, we will simply use the symbol $\mathbf{G}$ for such a generator. In the Markovian case the matrix $\overline{\mathbf{K}}(\nu)$ is equivalent to the generator of the stochastic process, in fact, $K_{x_i,x_j}(t) = g_{x_i,x_j}\delta(t)$ and its Laplace transform g_{x_i,x_j} is a time-independent rate.

In this section, we have seen that the stationary state of a semi-Markov process can be obtained as the stationary state of an appropriate *effective* Markov process, if we exclude pathological cases (such as WTDs with infinite mean, which can rule out such a stationary state). If we limit our attention only to this aspect, the efforts in studying the formalism of semi-Markov processes may not appear justified. In contrast to this, we will show in the next chapters that the fluctuations of certain quantities *cannot* be predicted on the basis of a Markovian description. This would be one of the take-home messages of this thesis. Before investigating in details such fluctuations, we consider another property for which the choice non-exponential WTD is relevant, i.e., the *reversibility* of a stochastic process.

1.4.2 Equilibrium and non-equilibrium

We now separate two important classes of stationary states, i.e., the *equilibrium* and *non-equilibrium* stationary states (for which the acronym NESSs is also used). Let us focus on the long-time limit of a semi-Markov process, where the probability of having a

certain state is time invariant and is given by equation (1.97). Recall that, in this limit, equations (1.92) and (1.94) are satisfied for each x_i. In other words, this means that total number of jumps into a state exactly balances the total number of jumps out of a state. For each pair (x_i, x_j), we define a probability *flow* as

$$\mathsf{j}_{x_i,x_j} = P^*_{x_j} \int_0^\infty K_{x_i,x_j}(t)\,\mathrm{d}t - P^*_{x_i} \int_0^\infty K_{x_j,x_i}(t)\,\mathrm{d}t \tag{1.98}$$

$$= P^*_{x_j} \frac{\mathsf{P}_{x_i,x_j}}{\mathsf{A}_{x_j}} - P^*_{x_i} \frac{\mathsf{P}_{x_j,x_i}}{\mathsf{A}_{x_i}}. \tag{1.99}$$

When the flow is zero for each $x_i, x_j \in \mathcal{S}$, we say that the stationary state satisfies the *detailed balance* condition

$$P^*_{x_j} \frac{\mathsf{P}_{x_i,x_j}}{\mathsf{A}_{x_j}} - P^*_{x_i} \frac{\mathsf{P}_{x_j,x_i}}{\mathsf{A}_{x_i}} = 0. \tag{1.100}$$

This definition comprises both discrete and continuous-time Markov processes; in the discrete time case we have that the waiting times A_{x_i} are constant in x_i, while in continuous-time memory-less processes $\mathsf{P}_{x_i,x_j}/\mathsf{A}_{x_j}$ is just the constant transition rate g_{x_i,x_j}. Equations (1.98) and (1.100) imply that a semi-Markov process satisfy detailed balance if and only if its effective Markov process satisfies detailed balance [55, 67].

Clearly, the stationarity does not require detailed balance, the latter being only a sufficient condition for the former. However, assuming this second condition is rather convenient, as it allows us to write immediately a recursive solution for the stationary state,

$$P^*_1 = \frac{\mathsf{P}_{1,0}}{\mathsf{A}_0}\frac{\mathsf{A}_1}{\mathsf{P}_{0,1}} P^*_0, \quad P^*_2 = \frac{\mathsf{P}_{2,1}}{\mathsf{A}_1}\frac{\mathsf{A}_2}{\mathsf{P}_{1,2}} P^*_1, \quad \ldots, \quad P^*_n = \frac{\mathsf{P}_{n,n-1}}{\mathsf{A}_{n-1}}\frac{\mathsf{A}_n}{\mathsf{P}_{n-1,n}} P^*_{n-1}, \tag{1.101}$$

in other words, $P^*_n = \prod_{i=1}^n \mathsf{P}_{i,i-1}\mathsf{A}_{i-1}/(\mathsf{P}_{i-1,i}\mathsf{A}_i) P^*_0$. This can be used to define a *free energy* difference $A_{x_i} - A_{x_j}$ between a pair of states, which is an additive quantity and determines their relative stationary probability as follows

$$\frac{P^*_{x_i}}{P^*_{x_j}} = \frac{\mathsf{P}_{x_i,x_j}\mathsf{A}_{x_i}}{\mathsf{P}_{x_j,x_i}\mathsf{A}_{x_j}} = \exp[-(A_{x_i} - A_{x_j})]. \tag{1.102}$$

Another way to write down equation (1.100) is

$$\sqrt{\frac{P^*_{x_j}}{P^*_{x_i}}}\,\frac{\mathsf{P}_{x_i,x_j}}{\mathsf{A}_{x_j}} = \sqrt{\frac{P^*_{x_i}}{P^*_{x_j}}}\,\frac{\mathsf{P}_{x_j,x_i}}{\mathsf{A}_{x_i}} =: [\mathbf{G}^*_s]_{x_i,x_j}\,, \tag{1.103}$$

which means that it is possible to define a symmetric matrix

$$\mathbf{G}_s^* = \mathbf{P}^* \, \mathbf{G}^* \, (\mathbf{P}^*)^{-1} \,, \tag{1.104}$$

where $\mathbf{P}^*$ is diagonal with entries $[\mathbf{P}^*]_{x_i,x_i} = \sqrt{P^*_{x_i}}$. This in turns implies another necessary condition for the detailed balance, i.e., the eigenvalues of the associated generator $\mathbf{G}^*$ are real.

We now mention a condition even stronger than detailed balance. In order to do so in general terms, we first need to introduce the following notion. The *reverse* of a stochastic process $X(t)$ is the stochastic process $X(\tau - t)$ for some given $\tau \in \mathcal{T}$, and a stochastic process $X(t)$ is *reversible* if

$$\begin{aligned} &\text{Prob}\{X(t_0) = x_0, X(t_1) = x_1, \ldots, X(t_n) = x_n\} \\ &\qquad = \text{Prob}\{X(\tau - t_0) = x_0, X(\tau - t_1) = x_1, \ldots, X(\tau - t_n) = x_n\}. \end{aligned} \tag{1.105}$$

for every $\tau \in \mathcal{T}$ and $t_0 \leq t_1 \leq \ldots \leq t_n$. In words this means that the probability of observing a sample path of $X(t)$ is equal to the probability of observing a sample path of the reverse process $X(\tau - t)$, indexed by t. If equation (1.105) is not satisfied, the process is said to be *irreversible*. These definitions are equally valid in discrete and continuous time. Roughly speaking, this condition is in general stronger than equation (1.100) as it requires that not only the number of jumps, but also the durations of the visits are balanced. While in Markov processes the detailed balance condition is necessary and sufficient condition for the reversibility of its stationary states [48], in semi-Markov processes the DTI is also needed, as proved by M. K. Chari [7] and evoked by following argument. Using equation (1.43) we write the probability density of observing a history $w(t)$ as

$$\begin{aligned} \varrho[w(t)] = P^*_{x_0} \frac{1}{\theta_{x_0}} \mathsf{P}_{x_1,x_0} \mathsf{P}_{x_2,x_1} \int_{t_1-t_0}^{\infty} f_{x_1,x_0}(t) \, \mathrm{d}t \, f_{x_2,x_1}(t_2 - t_1) \times \\ \ldots \times \mathsf{P}_{x_{n-1},x_{n-2}} \mathsf{P}_{x_n,x_{n-1}} f_{x_{n-2},x_{n-1}}(t_{n-2} - t_{n-1}) f_{x_{n-1},x_n}(t_{n-1} - t_n) \phi_{x_n}(t - t_n), \end{aligned} \tag{1.106}$$

while the probability density to observe the reverse history is

$$\begin{aligned} \varrho[w_R(t)] = P^*_{x_n} \frac{1}{\theta_{x_n}} \mathsf{P}_{x_{n-1},x_n} \mathsf{P}_{x_{n-2},x_{n-1}} \int_{t_n-t_{n-1}}^{\infty} f_{x_{n-1},x_n}(t) \, \mathrm{d}t \, f_{x_{n-2},x_{n-1}}(t_{n-1} - t_{n-2}) \times \\ \ldots \times \mathsf{P}_{x_2,x_1} \mathsf{P}_{x_1,x_0} f_{x_1,x_2}(t_2 - t_1) f_{x_0,x_1}(t_1 - t_0) \phi_{x_0}(t_1 - t_0). \end{aligned} \tag{1.107}$$

Now if we use the detailed balance $P^*_{x_j} = P^*_{x_i} \mathsf{P}_{x_j,x_i} / \mathsf{A}_{x_i}$ recursively and assume the DTI in the form of $f_{x_i}(t) = f_{x_j,x_i}(t)$, then the r.h.s. of equation (1.106) is equal to the

r.h.s. of equation (1.107), i.e., $\varrho[w(t)] = \varrho[w_R(t)]$. Vice-versa, suppose that the process is reversible. Then, $\varrho[w(t)] = \varrho[w_R(t)]$ for any finite history $w(t)$. In particular choosing $n = 2$ and equating (1.106) and (1.106), we get

$$P^*_{x_0} \frac{\mathsf{P}_{x_1,x_0}}{\theta_{x_0}} \int_{t_1-t_0}^{\infty} f_{x_1,x_0}(t)\,\mathrm{d}t\,\phi_{x_1}(t\,-\,t_1) \;=\; P^*_{x_1} \frac{\mathsf{P}_{x_0,x_1}}{\theta_{x_1}} \int_{t-t_1}^{\infty} f_{x_0,x_1}(t)\,\mathrm{d}t\,\phi_{x_0}(t_1\,-\,t_0), \tag{1.108}$$

for all $x_0, x_1 \in \mathcal{S}$ and $t_0 \leq t_1 \leq t$. In the limit as $t \to t_0$, i.e., $(t_1 - t_0) \to 0$ and $(t - t_1) \to 0$, this equation gives the detailed balance $P^*_{x_0}\mathsf{P}_{x_1,x_0}/\mathsf{A}_{x_0} = P^*_{x_1}\mathsf{P}_{x_0,x_1}/\mathsf{A}_{x_1}$. Then we can cancel these factors out of the equation (1.108). After, we are still allowed to take the limit as $t \to t_1$ to obtain

$$\int_{t_1-t_0}^{\infty} f_{x_1,x_0}(t)\,\mathrm{d}t = \phi_{x_0}(t_1 - t_0), \tag{1.109}$$

which implies that $f_{x_1,x_0}(t)$, is independent of x_1, hence the DTI.

A convenient way to measure how probable a sample path $w(t)$ is compared to its time-reversal $w_R(t)$ is to define the entropy production

$$S_t[w(t)] = \ln \frac{\varrho[w(t)]}{\varrho[w_R(t)]}, \tag{1.110}$$

which we recognise as a time-extensive observable due to the logarithm. More generically, we can characterise a non-equilibrium system by means of diverse time-extensive functionals $J[w(t)]$ (see, e.g., references [68, 69]), which can be used as measures of the "distance" from equilibrium [54], at least in Markov processes. Most of the literature in non-equilibrium statistical mechanics, deals with functionals that can be written as the sum

$$\boxed{\begin{array}{c}\text{Type A observable}\\ \mathrm{A}[w(t)] = \displaystyle\sum_{i=0}^{n-1} \theta_{x_{i+1},x_i}\end{array}} \tag{1.111}$$

of elementary contributions corresponding to configuration changes. These quantities are referred to as type A functionals in the relevant literature, e.g., in [24, 45], and may represent physical observables integrated over the observation time, e.g., the dynamical activity in a glassy system [24], the moles of metabolites produced in a biochemical pathway [9, 63], the customers served in a queuing network [61], the flow in interacting particle systems (IPSs) [14], or certain quantities in stochastic thermodynamics [36, 59]. Notably, D. Andrieux and P. Gaspard [3] proved that for semi-Markov processes with

DTI, the following fluctuation symmetry is satisfied

$$\frac{\text{Prob}\{\text{A}[w(t)] = a\}}{\text{Prob}\{\text{A}[w(t)] = -a\}} \sim \exp\left(Ea\right), \tag{1.112}$$

where E is a field conjugated to A.

Alternatively it is possible to consider "static" contributions that do not depend on the arrival state x_i, and define type B functionals

$$\boxed{\begin{array}{c}\text{Type B observable}\\ \text{B}[w(t)] = \displaystyle\sum_{i=0}^{n-1} \theta_{x_i}(t_{i+1} - t_i)\end{array}} \tag{1.113}$$

Examples of functionals of this type will be considered in more detail only in chapter 5. For the same process, functionals of type A and type B are related [24, 45].

Hereafter, we will refer to both type A and type B observables as *time-integrated currents* $J[w(t)]$ and to their empirical average $J[w(t)]/T$ simply as *currents*. However, with the exception of chapter 5, we will mainly deal with type A observables.

1.5 Beyond semi-Markov processes

It is possible to define stochastic processes more general than the semi-Markov process, where the WTDs

$$\text{Prob}\{\text{the jump is } x_{i-1} \to x_i \text{ and the transition is at age } \tau\} \tag{1.114}$$

depend parametrically not only on the configuration x_{i-1}, as in the semi-Markov processes, but also on events occurred during any instant in the past history. We make such a parametric dependence explicit by using $\psi_{x_i,x_{i-1}}[\tau; w(t)]$ to denote the probability (1.114), where t is the time at which the last transition occurred and τ is the age.

Processes of this type appear naturally in compounded complex systems, where temporal correlations are due to the joint effect of the mechanisms responsible for the single-particle jumps. Another way to define a non-Markov process is to assign WTDs that depend on the whole trajectory through the current $J[w(t)]$; as an example of models in this class, we will study the IPS of R. J. Harris [35] in section 4.4.2 of chapter 4.

For all such cases, the probability density (1.115) of observing a sample path $w(t)$ is

replaced by

$$\varrho[w(t)] = \phi_{x_n}[t - t_n; w(t_n)]\psi_{x_n,x_{n-1}}[t_n - t_{n-1}; w(t_{n-1})]\ldots \\ \times \psi_{x_1,x_0}[t_1 - t_0; w(t_0)]P_{x_0}(t_0), \quad (1.115)$$

where the WTD $\psi_{x_n,x_{n-1}}[t_n - t_{n-1}; w(t_{n-1})]$ that a transition from x_{n-1} to x_n occurs during the infinitesimal interval $[t_n, t_n + \mathrm{d}t)$ now depends explicitly on the history until t_{n-1}. The non-Markov nature of equation (1.115) prevents us deriving a general instant-wise Master equation in the form of (1.84). The implicit evolution law is then

$$P_x(t) = \sum_{n=0}^{\infty} \int_{t_0}^{t} \mathrm{d}t_1 \int_{t_1}^{t} \mathrm{d}t_2 \ldots \int_{t_{n-1}}^{t} \mathrm{d}t_n \sum_{x_0,x_1,\ldots,x_n} \delta_{x,x_n}\, \varrho[w(t)], \quad (1.116)$$

which gives the probability that a configuration $x \in \mathcal{S}$ is observed at $t > t_0$.

1.6 The large deviation approach

Non-equilibrium statistical physics is mostly concerned with non-zero currents, whose comprehension still is one of the greatest challenges of contemporary physics [69]. The typical values of such currents can be obtained by the knowledge of the stationary state $|P^*\rangle$. However, currents fluctuate in time, and to study such fluctuations we resort to the *large deviation theory*. The analysis of rare fluctuations is interesting for both applied and pure science: in applications, it is important to predict how likely it is to have rare but distressing events, while fundamental research also unveiled symmetries such as the one in equation (1.112), which are known as Fluctuation Theorems (as reviewed, e.g., in references [36, 59]).

In this section, we present a short introduction to the *large deviation theory* for currents. Following reference [65], we formulate the theory only for observables extensive[7] in the variable t. We require the probability density $\mathcal{P}(j,t)$ that the observable $J[w(t)]/T$, i.e., the total current divided by the observation time $T = t - t_0$, assumes a value j at time t. Such a time-averaged current is doomed to converge to its expectation value, denoted by $\langle j \rangle$, in the limit $t \to \infty$ (for finite t_0), as stated by the Central Limit Theorem. We say that the current obeys a large deviation principle if the limit

$$\hat{e}(j) = \lim_{t\to\infty} -\frac{\ln \mathcal{P}(j,t)}{t} \quad (1.117)$$

[7] We mention that large deviation results exist also for sub-extensive or super-extensive observables [13, 17]

exists, is finite and non-zero for some j different from $\langle j \rangle$. The limit (1.117) can be interpreted as the exponential decay rate of $\mathcal{P}(j,t)$ with respect to t and is then referred to as the *rate function*. It provides us with information about the limiting behaviour of the current. Using the symbol $\sim$ to indicate the asymptotic equivalence, the large deviation principle can hence be expressed as

$$\mathcal{P}(j,t) \sim \mathrm{e}^{-t\hat{e}(j)}. \tag{1.118}$$

The direct evaluation of $\hat{e}(j)$ is not straightforward for many systems of interest. In this thesis, we compute the rate function following a standard strategy which consists of some small intermediate steps. We first introduce the moment generating function of $J[w(t)]$,

$$Z(s,t) = \int \mathrm{e}^{-sJ[w(t)]}\, \varrho[w(t)]\, \mathrm{d}w(t), \tag{1.119}$$

and the scaled cumulant generating function (SCGF),

$$e(s) = -\lim_{t\to\infty} \frac{1}{t} \ln Z(s,t), \tag{1.120}$$

which is always concave in s and represents an intensive field conjugated to J. If equation (1.120) is differentiable, then the Gärtner–Ellis theorem [65] ensures that we can compute the rate function (1.117) as the *Legendre–Fenchel* (LF) *transform* of $e(s)$, that is,

$$\hat{e}(j) = \sup_s \{e(s) - s\, j\}. \tag{1.121}$$

Clearly, such a strategy has a limitation. In fact, the Legendre–Fenchel transform of a concave function, such as equation (1.120), is always convex, while it is possible to find random variables with a non-convex rate function [65]. Hence, care is needed when non-differentiable SCGFs are found.

We also mention that, according to the Varadhan's theorem (see again, for example, reference [65]), if the large deviation principle (1.118) is satisfied, then the SCGF (1.120) is given by the inverse transform

$$e(s) = \inf_j \{\hat{e}(j) + s\, j\}, \tag{1.122}$$

which is the inverse relation of (1.121) when $\hat{e}(j)$ is convex. Typically, the right hand sides of equations (1.121) and (1.122) are both referred to as the Legendre–Fenchel transforms of $e(s)$ and $\hat{e}(j)$, respectively, the former being used for concave functions, the latter for convex functions.

Now, we consider the special case where $e(s)$ is strictly concave, i.e., it does not contain any linear branch. Then equation (1.121) reduces to

$$\hat{e}(j) = e(s^*) - s^* j, \tag{1.123}$$

where s^* is the *unique* solution in s of

$$\partial e(s)/\partial s = j, \tag{1.124}$$

and the LF transform reduces to the *Legendre transform* of $e(s)$. The strict concavity implies that the function $f(s) = \partial e(s)/\partial s$ is monotonically decreasing, hence the inverse function f^{-1} exists. Applying such a function to both sides of equation (1.124), we get an implicit expression for s, in terms of j, namely $f^{-1}(j) = s$. It is convenient to define a new symbol s for such an inverse function to get

$$\mathrm{s}(j) = s. \tag{1.125}$$

It is worth stressing that, according to equation (1.124), the derivative of the original function $e(s)$ is the argument of the function $\hat{e}(j)$ obtained after a Legendre transform. We now show that the derivative of the Legendre transformed function is the argument of the original function. Using equation (1.125), we have that

$$\hat{e}(j) = e[\mathrm{s}(j)] - j\,\mathrm{s}(j), \tag{1.126}$$

hence

$$\frac{\partial \hat{e}(j)}{\partial j} = \frac{\partial e(\mathrm{s})}{\partial \mathrm{s}} \frac{\partial \mathrm{s}(j)}{\partial j} - \mathrm{s}(j) - \frac{\partial \mathrm{s}(j)}{\partial j}\, j, \tag{1.127}$$

Here, by virtue of equation (1.124), the first and third term of the r.h.s. cancel to leave $\partial \hat{e}(j)/\partial j = -\mathrm{s}(j)$, a relation analogous to equation (1.124). Therefore, when $e(s)$ is differentiable, its slope has a one-to-one correspondence with the slope of $\hat{e}(j)$, a property referred to as the *duality* of the Legendre transform. The graphical construction of $\hat{e}(j)$ from $e(s)$ is illustrated in figure 1.2.

Conversely, when $e(s)$ is not strictly concave, the LF transform maps all the points in the branch with constant slope $\partial e(s)/\partial s$ to the same point in $\hat{e}(j)$, as illustrated in figure 1.3. As a consequence, $\hat{e}(j)$ can be non-differentiable at such a point. It is worth noting that if the green linear branch in the function of figure 1.3(left) is replaced with a convex branch (with values less than or equal to those of the linear branch), then the LF transform remains identical to that of figure 1.3(right). This suggests that applying an LT transform to a non-strictly concave function does not guarantee that all the information on the original function is preserved. Such loss of information is physically relevant in

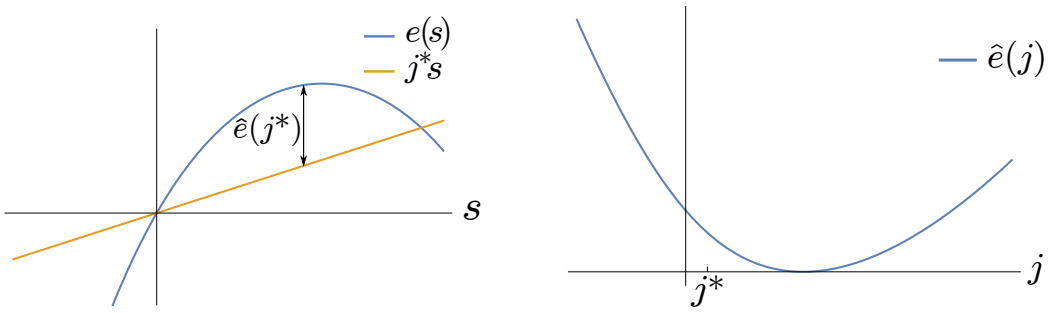

Figure 1.2: Graphical construction of the LF transform $\hat{e}(j)$ of a strictly concave function $e(s)$. For a given value of j^* there is only one value of $\sup_s\{e(s) - sj^*\}$ (left figure). This gives a value of the convex function $\hat{e}(j)$ (right figure). The original function $e(s)$ can be obtained applying the inverse transform (1.122). In this case the LF transform is equivalent to the simpler Legendre transform, as in equations (1.123) and (1.125), which only encodes for the generic extremum and does not make distinction between the infimum and the supremum.

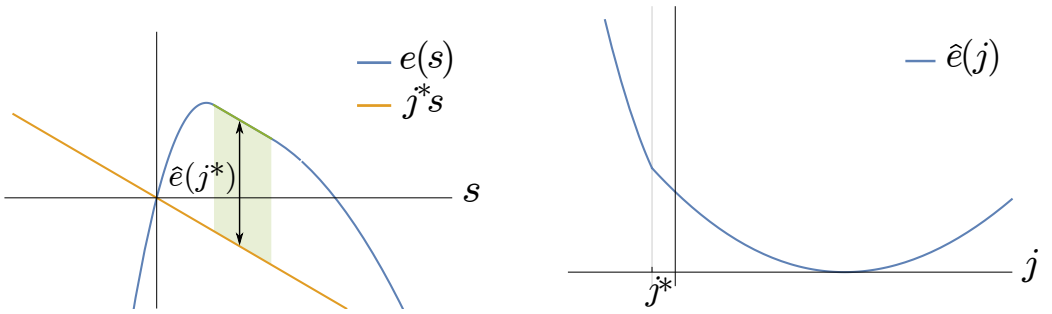

Figure 1.3: Graphical construction of the LF transform $\hat{e}(j)$ of a concave function $e(s)$ with a linear branch. All the points corresponding to the shaded area in the left figure are mapped to the point at j^* in the right figure. The values of $\hat{e}(j)$ for $j < j^*$ ($j > j^*$) depend on the values of $e(s)$ at the right (left) side of the shaded area.

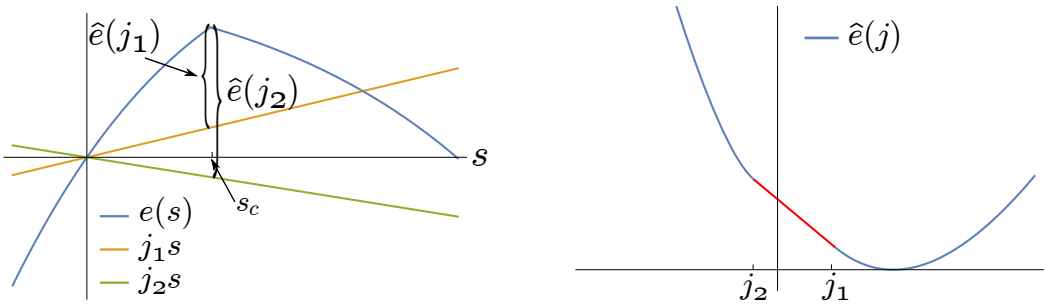

Figure 1.4: Construction of the LF transform $\hat{e}(j)$ of a strictly concave function $e(s)$ with a non-differentiable point s_c. The value $e(s_c) - js_c$ determines the values of $\hat{e}(j)$ along an interval $j_1 < j < j_2$ (red line in the right figure). As suggested by equation (1.124), the boundary values j_1 and j_2 are given by the left and right derivative of $e(s)$ at s_c, respectively.

the following scenario. Let us consider the LF transforms of non-differentiable functions. The definition of the Legendre transform is, of course, inadequate for this case; on the contrary, a LF transform can be easily computed even on non-differentiable points, see the example on figure 1.4(left). Such a picture shows that the non-differentiable point of $e(s)$ contributes a continuous set of values in $\hat{e}(j)$ (red branch in figure 1.4(right)) whose values varies linearly with j (which determines the slope of the orange and green line in figure 1.4(left)). We stress that, in such a case, the Gärtner–Ellis theorem does not apply and it is not guaranteed that $\hat{e}(j)$ is the correct rate function: in presence of non-differentiable SCGFs, the LF transform only provides the *convex hull* of the true rate function, which can hide a non-convex branch. In this thesis, we will encounter some non-differentiable SCGFs but the existence of non-convex rate function has been ruled out by numerical tests. The mathematically inclined reader is referred to reference [58] for a detailed and rigorous approach to convex analysis. Details about the mechanisms that render or hide a non-convex rate function, in the context of Statistical Mechanics, are in references [64, 66].

The SCGF of a type A current, in Markov processes, has a rather compact representation. We work on a joint configuration–current space, defined as the Cartesian product $\mathcal{S} \otimes \mathcal{J}$, where $\mathcal{J}$ is the set of all the values that the integrated current $J[w(t)]$ can assume. We also represent the random variable $J[w(t)]$ as a diagonal operator $\mathbf{J}$, whose diagonal elements are the set of possible values that the integrated current can assume. The generic element of this space is the column vector $|P'(t)\rangle = \sum_{x_i,\mathtt{J}} P_{x_i,\mathtt{J}}(t)|e_{x_i,\mathtt{J}}\rangle$ which represents the joint probability mass distribution that the system has a certain configuration and current; $|e_{x_i,\mathtt{J}}\rangle$ is the generic element of a natural basis for $\mathcal{S} \otimes \mathcal{J}$. Hence, if we begin the observation at the initial time $t_0 = 0$, the moment generating function can

be written as

$$Z(s,t) = \langle 1| \left(\mathbb{1} \otimes \mathrm{e}^{-s\mathbf{J}}\right) |P'(t)\rangle, \tag{1.128}$$

The vector $|P'(t)\rangle$ can be obtained from an initial state $|P'(0)\rangle$ with zero total current through $|P'(t)\rangle = \mathrm{e}^{\mathbf{G}'t}|P'(0)\rangle$, where $\mathrm{e}^{\mathbf{G}'t}$ is the time evolution operator in the joint configuration–current space and is tridiagonal with respect to the current sub-space. We diagonalise $\mathrm{e}^{\mathbf{G}'t}$ by means of a discrete Laplace transform, which in operator formalism reads $\left(\mathbb{1} \otimes \mathrm{e}^{-s\mathbf{J}}\right) \mathrm{e}^{\mathbf{G}'t} \left(\mathbb{1} \otimes \mathrm{e}^{s\mathbf{J}}\right) = \mathrm{e}^{\tilde{\mathbf{G}}t} \otimes \mathbb{1}$. This can be achieved by using the identity operator in the form $\mathrm{e}^{s\mathbf{J}}\mathrm{e}^{-s\mathbf{J}}$, to obtain $\langle 1| \left(\mathbb{1} \otimes \mathrm{e}^{-s\mathbf{J}}\right) \mathrm{e}^{\mathbf{G}'t} \left(\mathbb{1} \otimes \mathrm{e}^{s\mathbf{J}}\right) \left(\mathbb{1} \otimes \mathrm{e}^{-s\mathbf{J}}\right) |P'(0)\rangle$. Hence, we obtain an operator $\tilde{\mathbf{G}}$ which is regrettably non-stochastic but encodes for the statistics of the current without the need of the sub-space $\mathcal{J}$, i.e.,

$$Z(s,t) = \langle 1| \left(\mathrm{e}^{\tilde{\mathbf{G}}t} \otimes \mathbb{1}\right) \left(\mathbb{1} \otimes \mathrm{e}^{-s\mathbf{J}}\right) |P'(0)\rangle. \tag{1.129}$$

Since it has been assumed that at the beginning of the observation the current is zero, we can more compactly write

$$Z(s,t) = \langle 1|\mathrm{e}^{\tilde{\mathbf{G}}t}|P(0)\rangle, \tag{1.130}$$

where $|P(0)\rangle$ denotes the initial probability vector in the configuration subspace.

More explicitly, the generic row equation of the Master equation obeyed by the joint probability $P_{x_i,\mathsf{J}}(t)$ is

$$\frac{\mathrm{d}}{\mathrm{d}t} P_{x_i,\mathsf{J}}(t) = \sum_{x_j} \left[g_{x_i,x_j} P_{x_j,\mathsf{J}-\theta_{x_i,x_j}}(t) - g_{x_j,x_i} P_{x_i,\mathsf{J}}(t) \right]. \tag{1.131}$$

Multiplying by $\mathrm{e}^{-s\mathsf{J}}$ and summing over J both sides of equation (1.131), it is straightforward to show that the discrete Laplace transform

$$\tilde{P}_{x_i}(t) = \sum_{\mathsf{J}} \mathrm{e}^{-s\mathsf{J}} P_{x_i,\mathsf{J}}(t) \tag{1.132}$$

evolves according to

$$\frac{\mathrm{d}}{\mathrm{d}t} \tilde{P}_{x_i}(t) = \sum_{x_j} \left[g_{x_i,x_j} \mathrm{e}^{-s\theta_{x_i,x_j}} \tilde{P}_{x_j}(t) - g_{x_j,x_i} \tilde{P}_{x_i}(t) \right], \tag{1.133}$$

which can be thought of a Master equation with non-stochastic generator $\tilde{\mathbf{G}}$, obtained from $\mathbf{G}$ by multiplying by $\mathrm{e}^{-s\theta_{x_i,x_j}}$ the entries that contribute a term θ_{x_i,x_j} to the total current. The products $\mathrm{e}^{-s\theta_{x_i,x_j}} g_{x_i,x_j}$ are referred to as *biased* rates. Hereafter, we will refer to the tilded operator $\tilde{\mathbf{G}}$ as the a *biased* or *s-modified* generator. It is worth noting

that we can still apply the Perron–Frobenius theory of non-negative matrices[8] to $\mathrm{e}^{\tilde{\mathbf{G}}t}$ and expect that it has only one positive eigenvalue $\mathrm{e}^{\lambda_1 t}$ on the spectral radius. We also use the "tilded" notation to denote the generic column vector $|\tilde{P}(t)\rangle = \exp(\tilde{\mathbf{G}}t)|P(0)\rangle$. Let us denote by $|\tilde{P}_{\lambda_1}\rangle$ the right eigenvector of $\tilde{\mathbf{G}}$ associated with the largest eigenvalue λ_1; then, the long-time limit of the generating function is accessible through

$$Z(s,t) \sim \langle 1|\tilde{P}_{\lambda_1}\rangle\langle\tilde{P}_{\lambda_1}|P(0)\rangle\,\mathrm{e}^{\lambda_1 t}, \qquad t\to\infty, \tag{1.134}$$

as long as the pre-factors $\langle 1|\tilde{P}_{\lambda_1}\rangle$ and $\langle\tilde{P}_{\lambda_1}|P(0)\rangle$ are finite and a point spectrum exists. Under the same conditions, we can identify $-\lambda_1$ with the SCGF (1.120). Conversely, when one of the two pre-factors diverges, or when the leading eigenvalue is not known, we rely on the study of the integral representation of the moment generating function.

We finally mention that large deviation functionals can also be defined for observables extensive in a variable different than the time. For example, the equilibrium Statistical Mechanics is formulated in the limit as the system size approaches infinity with the total Energy being extensive in the size and thus taking the role of J. The analogues of s, $e(s)$ and $\hat{e}(j)$ are the inverse temperature β, the Helmholtz Free Energy scaled by the temperature $A\beta$ and the Entropy S, respectively [1, 16, 17, 65, 70]. This also justifies the use of "canonical" density and "partition function" when referring to $\mathrm{e}^{-sJ[w(t)]}\,\varrho[w(t)]$ and $Z(s,t)$, respectively.

1.7 Monte Carlo methods

1.7.1 Stochastic simulation algorithm

A stochastic simulation consists of the generation of sample trajectories of a stochastic process. In section 1.2 we defined a trajectory as the ordered collection of configuration–time pairs of equation (1.7). The general strategy used in the stochastic simulations is to decompose the problem of finding pairs of the type (t_i, x_i), and hence find the answer to two questions:

1) when is the next transition going to occur?
2) what configuration x_i will the system adopt after such a transitions?

For continuous-time Markov processes, this problem has been solved by J. L. Doob [15]; years later, D. T. Gillespie [28] obtained the same result by making use of physical argu-

[8] when the configuration space is finite.

ments. As the configuration x_i is determined by the knowledge of the configuration x_{i-1} before the last jump and the reactions that yield x_i, the question 2) can be replaced by:

2') what type of transition $x_{i-1} \to x_i$ will it be?

This makes explicit that, in a Markov process, the probability of having a certain configuration in the future depends on the configuration before the last jump.

The steps of a statistically correct stochastic-simulation algorithm can be obtained as follows. We divide and multiply equation (1.32) by $\sum_{x_j} g_{x_j,x_i}$. This gives representation (1.46), with

$$p_{x_j,x_i} = \frac{g_{x_j,x_i}}{\sum_{x_j} g_{x_j,x_i}} \tag{1.135}$$

and

$$\psi_{x_j}(\tau) = \left(\sum_{x_j} g_{x_j,x_i} \right) \exp\left(- \sum_{x_j} g_{x_j,x_i} \tau \right) \tag{1.136}$$

(notice that this case satisfies the DTI). The Gillespie–Doob algorithm for the generation of a sample path then reads:

1) Initialise the system to a configuration x_0 and a time t_0. Set a counter to $n = 1$.
2) Draw a value τ according to the density (1.136) and update the time to $t_n = t_{n-1} + \tau$.
3) Update the system configuration to x_n, with probability given by (1.135).
4) Update n to $n+1$ and repeat from 2) until t_n reaches the desired simulation time.

Widely used variants of the Doob-Gillespie scheme are the *tau-leaping* algorithm[9], which consists in combining many steps with exponentially-distributed waiting times into a single step encoding for a Poisson-distributed number of events [29], the exact *next reaction method* [27], wich uses only a single random number per simulation event, and the *random-sequential update* [56], which conversely samples more events to reject some of them[10]. More efficient, but still exact, versions of the stochastic-simulation algorithm for Markov processes are reviewed, e.g., in reference [30].

1.7.2 Population dynamics approach to the large deviations

According to the large deviation principle (1.118), the probability to observe a value of $J[w(t)]/T$ different from the mean value decreases exponentially with time. A consequence of this is that sample paths with current far from the mean are difficult to

[9] This is an approximate method.
[10] This is statistically correct, in agreement with the *thinning property* [34].

sample. For this reason advanced procedures are necessary to compute numerically the large deviation functionals.

In reference [25] a numerical procedure, which is referred to as the *cloning* or *population dynamics* method, was proposed as a general scheme for the evaluation of large deviation functionals of non-equilibrium Markov processes and demonstrated for a discrete-time Markov process. Such a procedure consists of assigning to each trajectory of the ensemble a population of identical clones consistent with a certain weight. As reviewed, e.g., in references [32, 33], the idea is not new. It seems to be born in the context of quantum physics and credited to E. Fermi [51], who suggested its use to evaluate the ground energy of a Schrödinger operator, see also reference [2] and the review of D. Ceperley and B. Alder [6]. Since then, it has also been extensively applied within equilibrium statistical physics [40, 46, 62] and proposed for applications in machine learning and probabilistic artificial intelligence [42]. This method gives direct access to the SCGF $e(s)$ and allow us to tune the parameter s as it would be possible in equilibrium thermodynamics, where we can tune, e.g., the temperature to obtain different values of the Helmholtz Free Energy.

The cloning approach to sample large deviation events in continuous-time Markov processes, which obey a Master equation for the form (1.14), has been demonstrated in reference [49] and has been widely used in literature, e.g., [4, 5, 18, 23, 24, 26, 43–45, 57]. Here we deal with the mathematics behind such a method, while a generalisation to non-Markovian processes and some situations that can lead to inefficiency or failure are discussed in chapter 4. Looking at the equation (1.119), we recognise that the modified generator $\tilde{\mathbf{G}}$ defines a prescription for the evolution in time of a vector $|\tilde{P}(t)\rangle$. The Master equation for this non-stochastic Hamiltonian reads

$$\frac{\mathrm{d}}{\mathrm{d}t}|\tilde{P}(t)\rangle = \tilde{\mathbf{G}}|\tilde{P}(t)\rangle. \tag{1.137}$$

Let us define $\mathbf{D}_{\tilde{\mathbf{G}}}$ and $\mathbf{D}_{\mathbf{G}}$ as diagonal matrices with entries

$$\left[\mathbf{D}_{\tilde{\mathbf{G}}}\right]_{x_i,x_i} = -\sum_{x_j} \mathrm{e}^{-s\theta_{x_j,x_i}} g_{x_j,x_i} =: \tilde{g}_{x_i}, \tag{1.138}$$

$$\left[\mathbf{D}_{\mathbf{G}}\right]_{x_i,x_i} = -\sum_{x_j} g_{x_j,x_i} =: g_{x_i} \tag{1.139}$$

respectively. By summing and subtracting $\mathbf{D}_{\tilde{\mathbf{G}}}$ on equation (1.137), we get

$$\frac{\mathrm{d}}{\mathrm{d}t}|\tilde{P}(t)\rangle = \tilde{\mathbf{G}}'|\tilde{P}(t)\rangle + \left(\mathbf{D}_{\tilde{\mathbf{G}}} - \mathbf{D}_{\mathbf{G}},\right)|\tilde{P}(t)\rangle. \tag{1.140}$$

where $\tilde{\mathbf{G}}'$ is a *stochastic* generator obtained from $\tilde{\mathbf{G}}$ after substituting its diagonal

with $\mathbf{D}_{\tilde{\mathbf{G}}}$. Equation (1.140), indeed, does *not* describe an ordinary stochastic evolution as $|\tilde{P}(t)\rangle$ does not satisfy the normalisation condition for probability vectors, i.e., $\langle 1|\tilde{P}(t)\rangle \neq \langle \tilde{P}(t)|\tilde{P}(t)\rangle \neq 1$, and does not conserve its norm as the system evolves over time. V. Lecomte and J. Tailleur [49] argued that the vector $|\tilde{P}(t)\rangle$ describes the stochastic dynamics of a set of trajectories evolving according to $\tilde{\mathbf{G}}'$ in time, while the weight associated to each trajectory expands or contracts as prescribed by $\left(\mathbf{D}_{\tilde{\mathbf{G}}} - \mathbf{D}_{\mathbf{G}}\right)$. A simulation algorithm follows straightforwardly:

1) Set up an ensemble of N clones and initialise each with a given time t_0, a random configuration x_0, and a counter $n = 0$. Set a variable C to zero. For each clone, draw a time τ of the next jump from the density $\left(\sum_{x_i} \tilde{g}_{x_i,x_0}\right) \exp\left(-\sum_{x_i} \tilde{g}_{x_i,x_0}\tau\right)$, and then choose the clone with the smallest value of $t = t_0 + \tau$.

2) For the chosen clone, update n to $n+1$, and the configuration x_{n-1} to x_n according to the probability mass
$$\tilde{g}_{x_{n+1},x_n} \Bigg/ \left(\sum_{x_i} \tilde{g}_{x_i,x_n}\right). \tag{1.141}$$

3) Generate a new waiting time τ for the updated clone according to
$$\left(\sum_{x_i} \tilde{g}_{x_i,x_n}\right) \cdot \exp\left(-\sum_{x_i} \tilde{g}_{x_i,x_n}\tau\right) \tag{1.142}$$
and increment the value of t to $t + \tau$.

4) **Cloning step.** Calculate a cloning factor
$$Y = \exp\left[\tau \cdot \sum_{x_i} (\tilde{g}_{x_i,x_n} - g_{x_i,x_n})\right]. \tag{1.143}$$
Discretise[11] the growth/decay rate to $y = \lfloor Y + u \rfloor$, where u is drawn from a uniform distribution on $[0, 1)$.

   1) If $y = 0$, prune the current clone.
   2) If $y > 0$, produce y copies of the current clones.

5) Increment C to $C + \ln[(N + Y - 1)/N]$. Choose the clone with the smallest t, and repeat from 2) until $t - t_0$ for the chosen clone reaches the desired simulation time T.

[11] The numerical caveats deriving from the discretisation of such a cloning rate are discussed in reference [38].

The SCGF is finally recovered as $-C/T$ for large T. A major problem is that this procedure implies a fast growth or decay of the total population, thus not being sustainable for many iterations. To tackle this issue the step 4) is supplemented with the following instruction, which sets the number of clones back to N:

4') 1) If $y = 0$, replace the pruned clone with another one, uniformly chosen among the remaining $N - 1$.
2) If $y > 0$, prune a number y of elements, uniformly chosen among the existing $N + y$.

This method has been reported to be inaccurate when the spectrum of the generator $\tilde{\mathbf{G}}$ is totally continuous [57] (this implies a vanished spectral gap and a slow convergence to $|\tilde{P}(t)\rangle\rangle$) and when the number of copies is close to the size of the ensemble [41].

Bibliography

[1] R. A. Alberty. Use of Legendre transforms in chemical thermodynamics (IUPAC Technical Report), 2001.

[2] J. B. Anderson. A random-walk simulation of the Schrödinger equation: H^+3. *J. Chem. Phys.*, 63(4):1499, 1975.

[3] D. Andrieux and P. Gaspard. The fluctuation theorem for currents in semi-Markov processes. *J. Stat. Mech.*, 2008(11):P11007, 2008.

[4] T. Bodineau, V. Lecomte, and C. Toninelli. Finite size scaling of the dynamical free-energy in a kinetically constrained model. *J. Stat. Phys.*, 147(1):1–17, 2012.

[5] M. Cavallaro, R. J. Mondragón, and R. J. Harris. Temporally correlated zero-range process with open boundaries: Steady state and fluctuations. *Phys. Rev. E*, 92(2): 022137, 2015.

[6] D. Ceperley and B. Alder. Quantum Monte Carlo. *Science*, 231(4738):555–560, 1986.

[7] M. K. Chari. On reversible semi-Markov processes. *Oper. Res. Lett.*, 15(3):157–161, 1994.

[8] A. C. C. Coolen and L. Holmberg. Principles of survival analysis. https://nms.kcl.ac.uk/ton.coolen/oup_survivalanalysis/, 2013. invited manuscript in preparation, for Oxford University Press, 2013.

[9] S. J. Court, B. Waclaw, and R. J. Allen. Lower glycolysis carries a higher flux than any biochemically possible alternative. *Nat. Commun.*, 6:8427, 2015.

[10] D. R. Cox. A use of complex probabilities in the theory of stochastic processes. *Mathematical Proceedings of the Cambridge Philosophical Society*, 51(02):313, 1955.

[11] D. R. Cox. *Renewal theory*. Chapman & Hall, London UK, 1967.

[12] D. R. Cox and H. D. Miller. *The Theory of Stochastic Processes*. Chapman & Hall, London UK, 1977.

[13] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Springer, Berlin Heidelberg Germany, 2010.

[14] B. Derrida. Non-equilibrium steady states: fluctuations and large deviations of the density and of the current. *J. Stat. Mech.*, 2007(07):P07023–P07023, 2007.

[15] J. L. Doob. Topics in the theory of Markoff chains. *Trans. Amer. Math. Soc.*, 52 (1):37–37, 1942.

[16] R. S. Ellis. An overview of the theory of large deviations and applications to statistical mechanics. *Scand. Actuar. J.*, 1995(1):97–142, 1995.

[17] R. S. Ellis. *Entropy, Large Deviations, and Statistical Mechanics*. Springer-Verlag, Berlin Heidelberg, Germany, 2006.

[18] C. Espigares, P. Garrido, and P. Hurtado. Dynamical phase transition for current statistics in a simple driven diffusive system. *Phys. Rev. E*, 87(3):032115, 2013.

[19] M. Esposito and K. Lindenberg. Continuous-time random walk for open systems: Fluctuation theorems and counting statistics. *Phys. Rev. E*, 77(5):051119, 2008.

[20] W. Feller. *An introduction to probability theory and its applications, Vol. 2*. Wiley, New York NY, 1957.

[21] O. Flomenbom and J. Klafter. Closed-Form Solutions for Continuous Time Random Walks on Finite Chains. *Phys. Rev. Lett.*, 95(9):098105, 2005.

[22] O. Flomenbom, K. Velonia, D. Loos, S. Masuo, M. Cotlet, Y. Engelborghs, J. Hofkens, A. E. Rowan, R. J. M. Nolte, M. Van der Auweraer, F. C. de Schryver, and J. Klafter. Stretched exponential decay and correlations in the catalytic activity of fluctuating single lipase molecules. *Proc. Nat. Ac. Sci. USA*, 102(7):2368–2372, 2005.

[23] J. P. Garrahan, R. L. Jack, V. Lecomte, E. Pitard, K. van Duijvendijk, and F. van Wijland. Dynamical first-order phase transition in kinetically constrained models of glasses. *Phys. Rev. Lett.*, 98(19):195702, 2007.

[24] J. P. Garrahan, R. L. Jack, V. Lecomte, E. Pitard, K. van Duijvendijk, and F. van Wijland. First-order dynamical phase transition in models of glasses: an approach based on ensembles of histories. *J. Phys. A: Math. Gen.*, 42(7):075007, 2009.

[25] C. Giardinà, J. Kurchan, and L. Peliti. Direct evaluation of large-deviation functions. *Phys. Rev. Lett.*, 96(12):120603, 2006.

[26] C. Giardinà, J. Kurchan, V. Lecomte, and J. Tailleur. Simulating rare events in dynamical processes. *J. Stat. Phys.*, 145(4):787–811, 2011.

[27] M. A. Gibson and J. Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem. A*, 104(9):1876–1889, 2000.

[28] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comp. Phys.*, 22(4):403–434, 1976.

[29] D. T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *J. Chem. Phys.*, 115(4):1716, 2001.

[30] D. T. Gillespie. Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.*,

58:35–55, 2007.

[31] L. Giuggioli, F. J. Sevilla, and V. M. Kenkre. A generalized master equation approach to modelling anomalous transport in animal movement. *J. Phys. A: Math. Gen.*, 42(43):434004, 2009.

[32] P. Grassberger. Go with the winners: a general Monte Carlo strategy. *Comp. Phys. Comm.*, 147(1-2):64–70, 2002.

[33] P. Grassberger and W. Nadler. 'Go with the winners' simulations. In *Computational Statistical Physics*, pages 169–190. Springer, Berlin Heidelberg Germany, Berlin, Heidelberg, 2002.

[34] G. Grimmett and D. Stirzaker. *Probability and Random Processes*. Oxford University Press, Oxford UK, 2001.

[35] R. J. Harris. Fluctuations in interacting particle systems with memory. *J. Stat. Mech.*, 2015(7):P07021, 2015.

[36] R. J. Harris and G. M. Schütz. Fluctuation theorems for stochastic dynamics. *J. Stat. Mech.*, 2007(07):P07020–P07020, 2007.

[37] J. W. Haus and K. W. Kehr. Diffusion in regular and disordered lattices. *Phys. Rep.*, 150(5-6):263–406, 1987.

[38] E. G. Hidalgo and V. Lecomte. Discreteness effects in population dynamics. *J. Phys. A: Math. Gen.*, 49(20):205002, 2016.

[39] T. Hoffmann, M. A. Porter, and R. Lambiotte. Generalized master equations for non-Poisson dynamics on networks. *Phys. Rev. E*, 86(4):046102, 2012.

[40] H.-P. Hsu and P. Grassberger. A review of Monte Carlo simulations of polymers with PERM. *J. Stat. Phys.*, 144(3):597–637, 2011.

[41] P. I. Hurtado and P. L. Garrido. Current fluctuations and statistics during a large deviation event in an exactly solvable transport model. *J. Stat. Mech.*, 2009(02): P02032, 2009.

[42] Y. Iba. Population Monte Carlo algorithms. *Trans. Jpn. Soc. Artif. Intell.*, 16(1): 279–286, 2001.

[43] R. L. Jack and P. Sollich. Large deviations and ensembles of trajectories in stochastic models. *Prog. Theo. Phys.*, 184:304–317, 2010.

[44] R. L. Jack and P. Sollich. Large deviations of the dynamical activity in the East model: analysing structure in biased trajectories. *J. Phys. A: Math. Gen.*, 47(1): 015003, 2014.

[45] R. L. Jack and P. Sollich. Effective interactions and large deviations in stochastic processes. *Eur. Phys. J. Spec. Top.*, 224(12):2351–2367, 2015.

[46] E. J. Janse van Rensburg. Monte Carlo methods for the self-avoiding walk. *J. Phys. A: Math. Gen.*, 42(32):323001, 2009.

[47] S. Karlin and J. L. McGregor. The differential equations of birth-and-death processes, and the Stieltjes moment problem. *T. Am. Math. Soc.*, 85(2):489–489, 1957.

[48] F. P. Kelly. *Reversibility and Stochastic Networks*. Cambridge University Press, Cambridge, UK, 2011.

[49] V. Lecomte and J. Tailleur. A numerical approach to large deviations in continuous time. *J. Stat. Mech.*, 2007(03):P03004–P03004, 2007.

[50] J. Masoliver, M. Montero, J. Perelló, and G. H. Weiss. The continuous time random walk formalism in financial markets. *J. Econ. Behav. & Organ.*, 61(4):577–598, 2006.

[51] N. Metropolis and S. Ulam. The Monte Carlo method. *J. Amer. Statist. Assoc.*, 44 (247):335–41, 1949.

[52] C. D. Meyer. *Matrix Analysis and Applied Linear Algebra*. SIAM, Philadelphia PA, 2000.

[53] E. W. Montroll and G. H. Weiss. Random Walks on Lattices. II. *J. Math. Phys.*, 6 (2):167, 1965.

[54] T. Platini. Measure of the violation of the detailed balance criterion: A possible definition of a "distance" from equilibrium. *Phys. Rev. E*, 83(1):011119, 2011.

[55] H. Qian and H. Wang. Continuous time random walks in closed and open single-molecule systems with microscopic reversibility. *EPL*, 76(1):15–21, 2007.

[56] N. Rajewsky, L. Santen, A. Schadschneider, and M. Schreckenberg. The Asymmetric Exclusion Process: comparison of update procedures. *J. Stat. Phys.*, 92(1-2):151–194.

[57] A. Rákos and R. J. Harris. On the range of validity of the fluctuation theorem for stochastic Markovian dynamics. *J. Stat. Mech.*, 2008(05):P05005, 2008.

[58] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princenton NJ, 1997.

[59] U. Seifert. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Rep. Progr. Phys.*, 75(12):126001, 2012.

[60] M. F. Shlesinger. Asymptotic solutions of continuous-time random walks. *J. Stat. Phys.*, 10(5):421–434, 1974.

[61] W. J. Stewart. *Probability, Markov Chains, Queues, and Simulation: The Mathematical Basis of Performance Modeling*. Princeton University Press, Princenton NJ, 2009.

[62] J. Tailleur and J. Kurchan. Probing rare physical trajectories with Lyapunov weighted dynamics. *Nature Phys.*, 3(3):203–207, 2007.

[63] M. Tomita and T. Nishioka. *Metabolomics: The Frontier of Systems Biology*. Springer, Tokyo Japan, 2006.

[64] H. Touchette. Simple spin models with non-concave entropies. *Am. J. Phys.*, 76(1): 26, 2008.

[65] H. Touchette. The large deviation approach to statistical mechanics. *Phys. Rep.*, 478(1-3):1–69, 2009.

[66] H. Touchette and C. Beck. Nonconcave Entropies in Multifractals and the Thermodynamic Formalism. *J. Stat. Phys.*, 125(2):455–471, 2006.

[67] H. Wang and H. Qian. On detailed balance and reversibility of semi-Markov processes and single-molecule enzyme kinetics. *J. Math. Phys.*, 48(1):1–15, 2007.

[68] R. K. P. Zia and B. Schmittmann. A possible classification of nonequilibrium steady

states. *J. Phys. A: Math. Gen.*, 39(24):L407–L413, 2006.

[69] R. K. P. Zia and B. Schmittmann. Probability currents as principal characteristics in the statistical mechanics of non-equilibrium steady states. *J. Stat. Mech.*, 2007 (07):P07012–P07012, 2007.

[70] R. K. P. Zia, E. F. Redish, and S. R. Mckay. Making sense of the Legendre transform. *Am. J. Phys.*, 77(7):614, 2009.

2 Temporally correlated zero-range process with open-boundaries: steady state

Contents

As a comprehensive theory of non-equilibrium phenomena does not exist, the study of analytically tractable toy models is an effective way to build it up. We begin our investigation by studying a model of interacting particles on lattice which we refer to as the *on-off zero-range process* (on-off ZRP). Generically, the zero-range paradigm requires that each lattice site can contain an arbitrarily large number of particles and that the particles leave the site after a random waiting time that depends only on the departure site configuration (thus encoding for on-site interactions). A modified ZRP, introduced in Hirschberg et al. [28], incorporates memory by means of an additional discrete phase variable for each lattice site, which has been referred to as a *clock* in the original literature. Specifically, in this chapter, we introduce and study the open-boundary version of such a model. We derive the exact NESS solution of the one-site system and a mean-field approximation for the general lattice. Both solutions correspond to that of a Markovian ZRP with effective interaction. We also explore the case where such NESS does not exists, a situation that is referred to as *congestion* or *condensation*.

This chapter is based on part of the results published in Cavallaro et al. [9] and is

organised as follows. The standard ZRP is inserted into the wider context of interacting-particle and queuing theory in section 2.1. The on-off ZRP with open-boundaries is introduced in section 2.2. The exact results for the one-site system and the mean-field approximation for the extended lattice are presented in sections 2.3 and 2.5 respectively, while the quantum Hamiltonian formalism to handle the dynamics on the extended lattice is discussed in section 2.4. The transition to the congested state is the topic of section 2.6.

2.1 The zero-range process and queuing networks

Modelling is an essential part of physics. A widespread approach to model real-world systems is to coarse-grain the physical space into a lattice, which is a discrete space with some boundary conditions. A typical lattice topology choice is the integer lattice, which is a subset Ω of the d-dimensional space $\mathbb{Z}^d$, $\mathbb{Z}$ denoting the set of integer numbers, although it possible to define different lattice structures. Also, periodic or open boundary conditions are typically specified. We assume that each site can be occupied by a certain number of particles, which can jump to another site according to the lattice topology and some dynamical rules. This defines an interacting-particle system (IPS) [34–36]. Formally, the state of the system is an element of the vector space $\mathbb{N}^{\Omega}$ and specifies the occupation number of each site.

IPSs feed a lot of very active research in applied probability [37, 38], and a comprehensive review is infeasible here. It is worth mentioning though that one of the most important models of IPSs is the *asymmetric simple exclusion processes* (ASEP). In such a model each lattice site can contain up to one particle, which attempts to leave the site after an exponentially distributed waiting time. Each static configuration is represented by an element of $\{0,1\}^{\Omega}$. Even in the simple one-dimensional lattice, the ASEP, with open or periodic boundaries, has been used to model real-life situations, such as surface growth and vehicular traffic; it also has interesting exact analytical results, see the details in, e.g., the reviews [6, 12, 15, 26, 39]. Also, modified ASEPs on one-dimensional lattice have been used to model the transport of kinesin and dynein along cytoskeleton microtubules [32], ribosome dynamics during protein translation [45], and molecule transport through microscopic channels [11].

The ZRP is another paradigmatic IPS, where each lattice site can contain an arbitrary positive number of particles. We are concerned with a one-dimensional chain lattice, see figure 2.1. The evolution proceeds in continuous time, i.e., transitions occur after a waiting time which is an exponentially distributed random variable. Specifically, in the

standard ZRP, a particle can hop to one of the neighbouring sites with rate proportional to μ_n, which depends only on the occupation number n of the departure site. Obviously, the departure rate from an empty site is given by $\mu_0 = 0$. The functional form of μ_n encodes for the interaction between particles, which occurs only on the departure site, hence the epithet "zero range". The special case $\mu_n = an$, where $a > 0$ is a constant, corresponds to free particles since in this case each particle leaves the site independently from the others. Other choices of μ_n correspond to attractive or repulsive inter-particle interaction if the n-dependence is sublinear or superlinear, respectively.

Models with zero-range interactions have proven to display complex collective behaviour whilst allowing analytical treatment [18]. In particular, the ZRP is well suited for theoretical analysis because the stationary probability distribution of a given configuration factorises and can be calculated exactly [46]. The stationary probability $P^*_{\{n_l\}_l}$ of finding the system in a state $\{n_l\}_l := (n_1, n_2, \dots n_L)$ with n_l particles on site l, $0 \leq l \leq L$, is given by a simple factorised form

$$P^*_{\{n_l\}_l} = \prod_{l=1}^{L} P^*_{n_l;l}, \tag{2.1}$$

where $P^*_{n_l;l}$ is the probability of finding the site l with n_l particles (this is also verified in presence of Poisson arrivals from the boundaries). The one-site marginals are determined by

$$P^*_{n;l} = \frac{{z_l}^n}{Z_l} \prod_{i=1}^{n} \mu_i^{-1}, \tag{2.2}$$

where z_l is a site-dependent fugacity (which is a function of the hopping rates) and the grand-canonical partition function $Z_l = \sum_{n=0}^{\infty} {z_l}^n \prod_{i=1}^{n} \mu_i^{-1}$ ensures normalisation [33]. It is worth noting that, for certain choices of μ_n, it is not possible for the sum in Z_l to converge for all z_l. The divergence of Z_l corresponds to the accumulation of particles on the site l and is referred to as congestion. Indeed, the infinite accumulation on one or more sites in an open system can be thought of as a kind of condensation phenomenon [10, 33]. Hence, in the following, we will also use the "condensation" terminology even for the single-site case. Condensation transitions far from equilibrium have been studied in physics [17], as well as in economics [7, 8], biology [19], network science [5, 13], and queueing theory [10]. Toy models, such as the ZRP, provide a theoretical foundation for understanding condensation in these systems. Exact results from the ZRP have also been used in models of vehicular traffic [31, 43], reptation in polymer physics [1], and transport and coalescence in granular systems [48].

The ZRP generalises to some extent the Jackson network [10, 14], which has been developed in Queuing Theory (a branch of the Operational Research community) [30, 47].

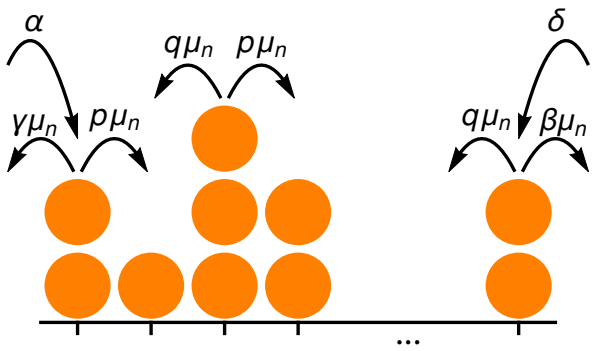

Figure 2.1: Zero-range process in one-dimensional lattice with open boundaries. A particle on a site with occupation number n leaves the site leftwards and rightwards with rate proportional to μ_n. New particles enter the lattice through the rightmost and leftmost sites.

Queuing networks are tools for the analysis of systems of customers that circulate among a certain number of nodes, waiting for and then receiving service on each node. They have also been proposed as tools for gene regulatory networks [2, 24] and epidemics [50]. Essentially, queuing networks can be safely thought of as IPSs, where the notion of particle is replaced by the one of "packet" or "customer". Such systems can also have open boundaries, where customers arrive from or leave to the outside world (after the service they asked for is completed); alternatively, closed networks do not receive new customers and the existing customers cycle indefinitely in the system. Typically, the stochastic dynamics at each node is characterised by the following quantities:

1. the arrival discipline A,
2. the service time policy B,
3. the number of servers C,
4. the node capacity N, i.e., the maximum occupation number.

According to the *Kendall notation* these are packed into an expression of the form A/B/C/N which defines a single queue. Some possible values for A and B are M for Markovian (i.e., exponentially distributed) inter-event times, Ph for phase-type distributed inter-event times, G for general policy, and D for deterministic or constant policy. The case with M/M/1/∞ corresponds to a ZRP with departure rate μ_n constant in n, while the case M/M/∞/∞ can be described by a ZRP with $\mu_n \propto n$. The ZRP is more general as it allows an arbitrary functional form for μ_n.

In networks of queues, the arrival discipline A cannot be specified a priori; however, a remarkable property of nodes of the type described above, when allowed to exchange customers/particles, is that each one of them behaves as if it were subjected to Poisson ar-

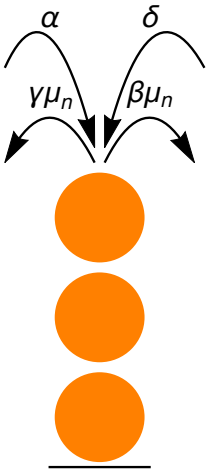

Figure 2.2: Single-site zero-range process. Particles enter the system with constant rate $\alpha + \delta$ while they leave with rate $\mu_n(\beta + \gamma)$, which depends on the occupation number n.

rivals, which in turns implies a factorised stationary solution of the form of equation (2.1) (which also holds with open boundaries, if the arrivals are Poisson). The validity of this statement is expressed by the Jackson theorem (see the reprint [30]), which applies to a wide class of queuing systems, including the BCMP networks [4] and the Gelenbe networks [20, 21, 25] (these are related to *spiked random neural networks* [22, 23]). One of the defining properties of such systems is that the node capacity N is infinite.

To illustrate the notions used in this and in the following chapter, we consider first a single-site lattice in contact with two boundaries (namely *left* and *right*), as shown in figure 2.2. Particles enter the site with constant rate, specifically the rate is α for particles injected from the left boundary and is δ for particles injected from the right boundary. As the configuration of the system is represented by the site occupation number, the probability of having the site with n particles at time t is written as $P_n(t)$. On-site particles leave the site rightwards or leftwards with rate $\mu_n\beta$ or $\mu_n\gamma$, respectively. The total extraction rate is $\mu_n(\beta + \gamma)$, which indeed depends only on the site configuration.

The Master equation can thus be written as a set of linear differential equations for $P_n(t)$,

$$\frac{\mathrm{d}}{\mathrm{d}t}P_n(t) = (\alpha + \delta)P_{n-1}(t) - (\alpha + \delta + (\beta + \gamma)\mu_n)P_n(t) + (\beta + \gamma)\mu_{n+1}P_{n+1}(t), \qquad (2.3)$$

when $n > 0$, with boundary conditions

$$\frac{\mathrm{d}}{\mathrm{d}t}P_0(t) = -(\alpha + \delta)P_0(t) + (\beta + \gamma)\mu_1 P_1(t), \qquad (2.4)$$

This system also represents an immigration–death process in continuous time. We compute the stationary state P_n^* as solution of $\frac{\mathrm{d}}{\mathrm{d}t}P_n(t) = 0$ for all $n \geq 0$, i.e., by solving the remaining difference equations

$$(\alpha+\delta)P_{n-1}^* - (\alpha+\delta+(\beta+\gamma)\mu_n)P_n^* + (\beta+\gamma)\mu_{n+1}P_{n+1}^* = 0, \tag{2.5}$$

for $n > 0$, with the initial condition

$$P_1^* = \frac{\alpha+\delta}{(\beta+\gamma)\mu_1}P_0^*. \tag{2.6}$$

The solution is

$$P_n^* = Z^{-1}\left(\frac{\alpha+\delta}{\beta+\gamma}\right)^n \prod_{i=1}^{n} {\mu_n}^{-1}, \tag{2.7}$$

where the factor $Z^{-1} = P_0^*$ is obtained imposing the condition $\sum_n P_n^* = 1$ and satisfies

$$Z = \sum_{n=0}^{\infty}\left(\frac{\alpha+\delta}{\beta+\gamma}\right)^n \prod_{i=1}^{n} {\mu_n}^{-1}. \tag{2.8}$$

This can be seen as a grand-canonical partition function with fugacity

$$z = \frac{\alpha+\delta}{\beta+\gamma}, \tag{2.9}$$

in analogy to the *equilibrium* Statistical Mechanics. In fact, for the one-site case, such an analogy is very strict, as the solution does not need the distinction between left and right boundaries (which is instead important for the next sections). The partition function Z allows us to express the mean occupation number as the logarithmic derivative

$$\langle n\rangle = \frac{\sum_{n=0}^{\infty} n z^n \prod_{i=1}^{n}\mu_i^{-1}}{\sum_{n=0}^{\infty} z^n \prod_{i=1}^{n}\mu_i^{-1}} = z\frac{\partial Z(z)}{\partial z}\frac{1}{Z(z)} = z\frac{\partial}{\partial z}\ln Z(z), \tag{2.10}$$

and, if we differentiate twice $\ln Z(z)$

$$\frac{\partial^2}{\partial z^2}\ln Z(z) = \frac{\partial^2 Z(z)}{\partial z^2}\frac{1}{Z(z)} - \left(\frac{\partial Z(z)}{\partial z}\frac{1}{Z(z)}\right)^2 = \frac{\langle n^2\rangle}{z^2} - \frac{\langle n\rangle^2}{z^2}, \tag{2.11}$$

we get the variance

$$\sigma^2 = z^2\frac{\partial^2}{\partial z^2}\ln Z(z). \tag{2.12}$$

The partition function also governs the range of validity of equation (2.7). In fact tuning the parameters α, β, γ and δ, as well as the interaction factor μ_n, it is possible for Z either to diverge or converge to zero. In such situations, the stationary solution (2.7) is not valid any more and we say that a *phase transition* occurs. A classical physical example has

been provided in the seminal work of C. N. Yang and T. D. Lee [53], stating that the zeros of the grand-canonical partition function in the asymptotic limit as the volume approaches infinity correspond to phase transitions (see also reference [6]). The poles in the asymptotic representations of the partition function also represent phase transitions, as discussed in reference [49]. For the one-site ZRP, the partition function (2.8) diverges when $\beta + \gamma \leq \alpha + \delta$, which means that the average occupation number is doomed to grow indefinitely and is infinite in the asymptotic limit as the time goes to infinity[1]. In the language of Queueing Theory the divergence of Z is referred to as *congestion*, while in physics literature, an analogue situation is better called *condensation*,

2.2 The on-off zero-range process

Several generalisations of the ZRP have been proposed, such as the "inclusion process" [27] and variants (some of which can display a phase described as "exploding" condensate [52]). We are interested in a modified zero-range process with non-Markovian dynamics, which has been introduced in Hirschberg et al. [28]. Its crucial new ingredient is that each site has an additional *clock/phase* variable τ and the particles cannot leave the site when the clock is set to zero, which corresponds to the off phase. The clock ticks and turns on with rate c and turns off with each particle arrival. This mechanism favours the accumulation of particles on a site. According to the zero-range dynamics, the particles interact only on site, but now have a different departure rate $\mu_{n,\tau}$. The additional variable τ takes into account events in particle configuration space that happened in the past and therefore introduces temporal correlations. Being an integer, τ defines distinct stages, as in the case 2) seen in section 1.3.1. This model has sparked interest as it displays, under certain conditions, a condensate with slow drift [28, 29]. Systems with distinct on and off phases are also of interest as models for intra-cellular ion-channels [40, 51] and for data traffic streams [41], as well as providing examples of stochastic processes with non-convex rate functions [16].

To the best of our knowledge, the ZRP with on-off dynamics has been studied only on ring topology, i.e., with periodic boundary conditions. In this and the following chapter we investigate the open-boundary version of the model, thus extending the work of Hirschberg et al. [28, 29]. We implement the same dynamics on an open chain with arbitrary hopping rates and boundary parameters, see figure 2.3. Particles are added and removed through the boundaries. On the leftmost lattice site (site 1), particles are injected with rate α and they are removed with rate $\gamma\mu_{n,\tau}$ which is non-zero only when the

[1] Notice that this is still represented by an *honest* stochastic process, as explained in section 1.2.

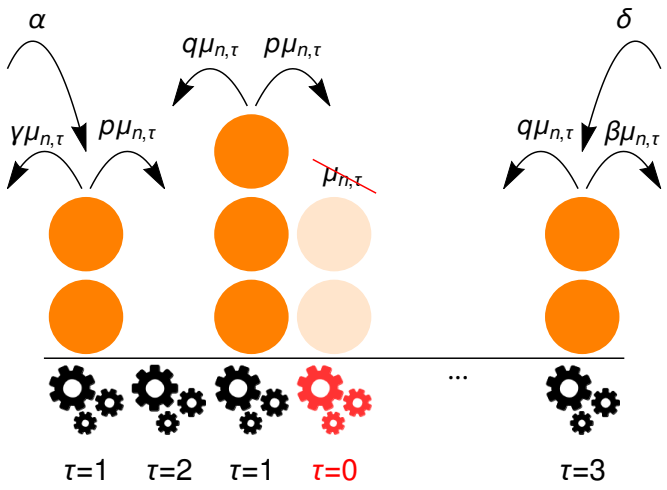

Figure 2.3: Non-Markovian ZRP on a one-dimensional lattice with open boundaries. Each site has a hidden variable τ, whose values are represented by the positions of a gear, which controls the departure rate. When τ assumes value zero no departure is possible and the corresponding state is referred to as off. This lock-down occurs in conjunction with the arrival of a particle.

phase of site 1 is different from $\tau = 0$. Similarly, on the rightmost site (site L) particles are removed with rate $\beta\mu_{n,\tau}$, according to the phase of site L, and are injected with rate δ. This situation corresponds to a bulk system in contact with two different reservoirs. In the bulk, particles jump to the left (right) with rate $q\mu_{n,\tau}$ ($p\mu_{n,\tau}$), which is again non-zero only when the phase of the departure site is not $\tau = 0$. The dynamics is sensitive to the specific rate values and we consider choices that induce a rightwards driving along the chain. In particular, it is worth making the distinction between a process with $(\gamma, \delta, q) = (0, 0, 0)$, which is referred to as totally asymmetric (TA), and a process where these equalities are not satisfied, generally referred to as partially asymmetric (PA).

Specifically, we will consider explicitly two forms for the interaction factor $\mu_{n,\tau}$, i.e., the case where $\mu_{n,\tau}$ is constant with respect to $n > 0$, which corresponds to an on-site attractive interaction between particles, and the case where $\mu_{n,\tau}$ is linear in n, which corresponds to no direct interaction between particles (excluding residual correlations due to the blockade mechanism).

The stationary state of the standard ZRP on an open chain has been extensively studied in Levine et al. [33]. In this case the particles are distributed along the system according to the product-form structure of equation (2.1) that implies no correlations between sites. In contrast, the on-off ZRP can generate more complex patterns, as shown in figure 2.4 for four sets of parameters. The clock-tick rate c plays a major role in these

patterns, the lower its value is, the more important the correlations are. Increasing the value of c, the system eventually becomes spatially uncorrelated. We now study in detail how the introduction of time correlations affects the stationary state.

2.3 Exact results for the one-site system

One of the most striking effects of the phase $\tau = 0$, which hereafter will be referred to as off phase, is to create correlations between the occupation numbers on each sites. Consequently there is no factorised steady state for the dynamics described in section 2.2. However, for the single-site system an exact solution is straightforward. First of all, as long as we are interested in the stationary state of the single site system, we do not need to distinguish between left and right boundary terms. The state is defined by two variables: the number of particles in the box n and a "clock" or "phase" variable τ, thus $(n, \tau) \in \mathcal{S}$. We focus on TA dynamics and consider a box which receives particles with rate α and ejects particles with rate $\beta\mu_{n,\tau}$, where $\mu_{n,\tau}$ is a function of the box state. The departure event is possible only when $\tau \neq 0$. Also, the dynamics includes the advance of the clock with rate c, and the reset to $\tau = 0$ when a particle arrives. If one defines $P_{-1,\tau}(t) = P_{n,-1}(t) = 0$, the following Master equation is valid for $\tau \geq 0$ and $n \geq 0$:

$$\frac{\mathrm{d}}{\mathrm{d}t}P_{n,\tau}(t) = cP_{n,\tau-1}(t) + \beta\mu_{n+1,\tau}P_{n+1,\tau}(t) + \delta_{\tau,0}\sum_{\tau'\geq 0}\alpha P_{n-1,\tau'}(t) - (c + \beta\mu_{n,\tau} + \alpha)P_{n,\tau}(t), \tag{2.13}$$

where $P_{n,\tau}(t)$ denotes the probability of finding the system with n particles and phase τ at time t and $\delta_{\tau,0}$ is a Kronecker delta. The first term on the right-hand side of (2.13) corresponds to a clock tick, the second term to the departure of a particle, the third term to the arrival of a particle and the fourth term to the respective escape events from the state (n, τ). The transition graph for this process is illustrated in figure 2.5.

As in references [28, 29], we choose to simplify the dependence of the jump rate on τ to $\mu_{n,\tau} = \mu_n$ when $\tau > 0$. Hereafter we specialise to this case, except when we explicitly refer to a general form for $\mu_{n,\tau}$. In this simplified case, it is more convenient to write the Master equation (2.13) in terms of $P_{n,\mathrm{ON}}(t) = \sum_{\tau>0} P_{n,\tau}(t)$ and $P_{n,\mathrm{OFF}}(t) = P_{n,0}(t)$,

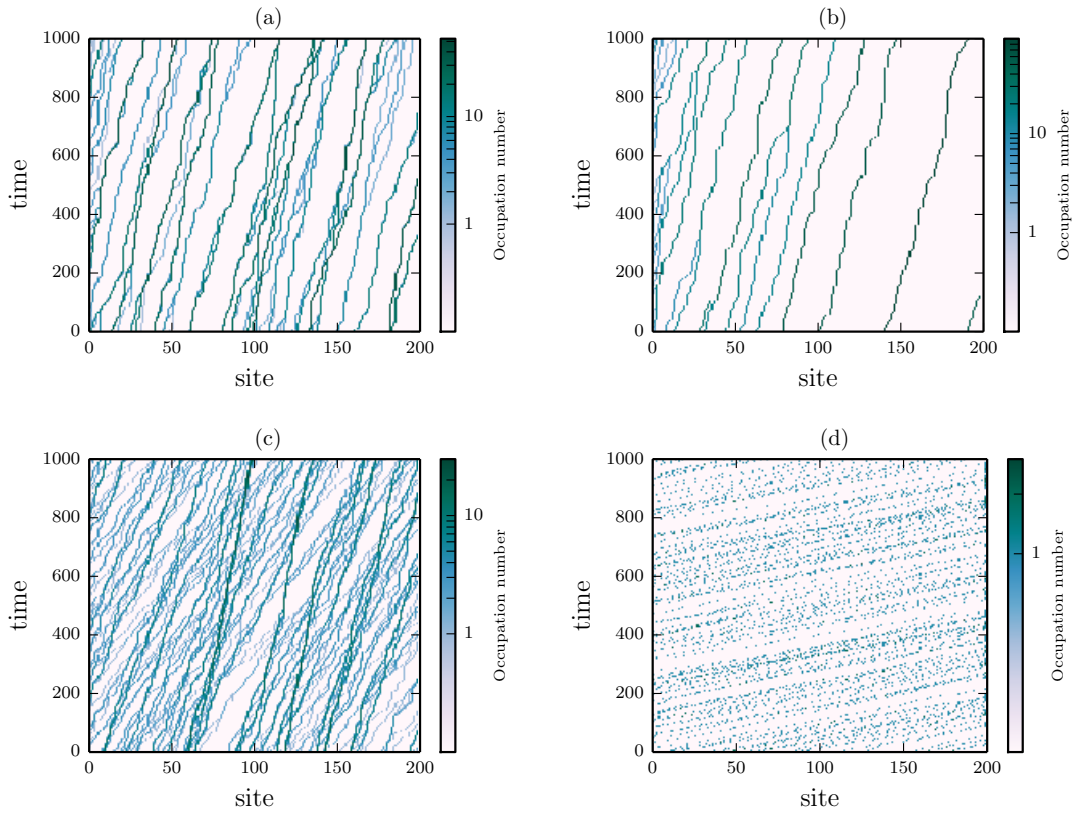

Figure 2.4: Monte Carlo time evolution of the occupation profile of the on-off ZRP on a one-dimensional lattice. (a) Rates $\mu_{n,\tau} = n$ if $\tau > 0$, $\mu_{n,\tau} = 0$ otherwise, and $(\alpha, \beta, \gamma, \delta, p, q, c) = (0.1, 0.2, 0, 0, 1, 0, 0.05)$. Only correlations due to the blockade mechanism are present. The particles organise in travelling clusters. Their speed is mainly governed by c. (b) Same parameters as the former case, except $(\beta, p) = (10^4, 10^4)$. The particles jump almost simultaneously to the next site as soon as the blockade is removed. Each particle cluster tends to occupy a single site. The drift proceeds with a rate $\simeq c$. (c) Same parameters as (a), except $\mu_{n,\tau} = 1$ if $\tau > 0$ and $n > 0$, $\mu_{n,\tau} = 0$ otherwise, and $c = 0.15$. As a result of the attractive inter-particle interaction, the clusters with more particles travel slower than the less populated ones. This mechanism enhances congestion. (d) Standard ZRP. Parameters are as in (a), except for the value of c, which can be assumed to be infinite in this case. The density profile is rather uniform.

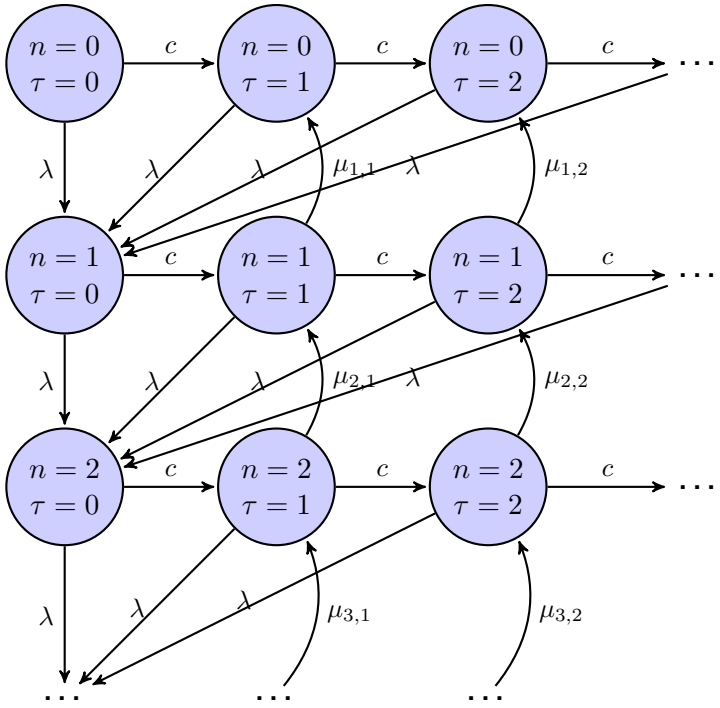

Figure 2.5: Markov graph of the one-site on-off ZRP, in the extended configuration–phase space.

i.e.,

$$\frac{\mathrm{d}}{\mathrm{d}t}P_{n,\mathrm{ON}(t)} = cP_{n,\mathrm{OFF}}(t) + \beta\mu_{n+1}P_{n+1,\mathrm{ON}}(t) - (\beta\mu_n + \alpha)P_{n,\mathrm{ON}}(t), \tag{2.14}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}P_{0,\mathrm{ON}}(t) = cP_{0,\mathrm{OFF}}(t) + \beta\mu_1 P_{1,\mathrm{ON}}(t) - \alpha P_{0,\mathrm{ON}}(t), \tag{2.15}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}P_{n,\mathrm{OFF}}(t) = \alpha P_{n-1,\mathrm{ON}}(t) + \alpha P_{n-1,\mathrm{OFF}}(t) - (c+\alpha)P_{n,\mathrm{OFF}}(t), \tag{2.16}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}P_{0,\mathrm{OFF}}(t) = -(c+\alpha)P_{0,\mathrm{OFF}}(t). \tag{2.17}$$

By equating the left-hand sides of equations (2.14)–(2.17) to zero, we find that the stationary distribution is given by

$$P_n^* = \frac{z^n}{Z_c}\prod_{i=1}^{n} w_{c,i}^{-1}, \tag{2.18}$$

$$P_{n,\mathrm{OFF}}^* = \frac{\beta\mu_n}{\alpha + c + \beta\mu_n}P_n^*, \tag{2.19}$$

$$P_{n,\mathrm{ON}}^* = \frac{(\alpha + c)}{\alpha + c + \beta\mu_n}P_n^*, \tag{2.20}$$

where $z = \alpha/\beta$, $w_{c,i} = \mu_i(\alpha + c)/(\alpha + c + \beta\mu_i)$, $Z_c = \sum_{n=0}^{\infty} z^n \prod_{i=1}^{n} w_{c,i}^{-1}$, and $P_n^* =$

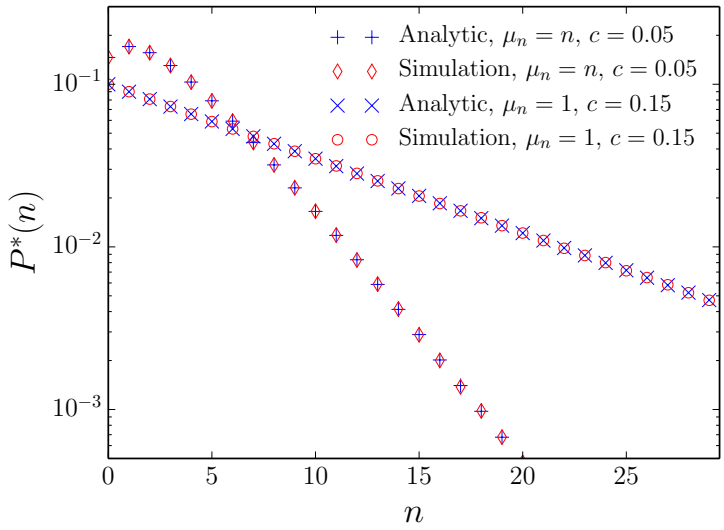

Figure 2.6: Occupation probability distribution of the one-site system for constant ($\mu_n = \mu, n > 0$) and linear ($\mu_n = n$) microscopic departure rate. The arrival and departure rate are $\alpha = 0.1$ and $\beta = 0.2$ respectively.

$P^*_{n,\mathrm{ON}} + P^*_{n,\mathrm{OFF}}$ by construction. We recognise the same stationary state (2.2) as the standard ZRP, with an effective departure rate $w_{c,n} = \mu_n P_{\mathrm{ON}|n}$, where $P_{\mathrm{ON}|n} = (\alpha + c)/(\alpha+c+\beta\mu_n)$ is the conditional probability of finding the site in the on state, given that there are n particles. For $c \to \infty$, the effective jump rate converges to the microscopic rate, i.e., $w_{c,n} \to \mu_n$. The stationary probability distribution of the occupation number is checked against Monte Carlo simulations in figure 2.6 for both constant and linear departure rates. Its tail is longer than the corresponding Markovian model ($c \to \infty$).

The equations (2.18)–(2.20) can be derived as follows. Summing equations (2.14) and (2.16), and imposing the stationarity condition, it follows that

$$\beta\mu_{n+1}P^*_{n+1,\mathrm{ON}} - \alpha P^*_n = \beta\mu_n P^*_{n,\mathrm{ON}} - \alpha P^*_{n-1}, \tag{2.21}$$

while the stationarity conditions on equations (2.15) and (2.17) imply the boundary conditions

$$\beta\mu_1 P^*_{1,\mathrm{ON}} - \alpha P^*_{0,\mathrm{ON}} = 0, \tag{2.22}$$

$$P_{0,\mathrm{OFF}} = 0, \tag{2.23}$$

which, together with (2.21), allow us to write the recursive relation

$$\beta\mu_{n+1}P^*_{n+1,\mathrm{ON}} = \alpha P^*_{n,\mathrm{ON}} + \alpha P^*_{n,\mathrm{OFF}}. \tag{2.24}$$

Using the stationarity condition on equation (2.16)

$$(\alpha + c)P^*_{n+1,\mathrm{OFF}} = \alpha P^*_{n,\mathrm{ON}} + \alpha P^*_{n,\mathrm{OFF}}, \tag{2.25}$$

we eliminate P^*_n from equations (2.24) and (2.25) and get

$$(\alpha + c)P^*_{n+1,\mathrm{OFF}} = \beta\mu_{n+1}P^*_{n+1,\mathrm{ON}}. \tag{2.26}$$

This, along with $P^*_n = P^*_{n,ON} + P^*_{n,OFF}$, immediately yields equations (2.19) and (2.20). The ratios $(\beta\mu_n)/(\alpha+c+\mu_n\beta)$ and $(\alpha+c)/(\alpha+c+\beta\mu_n)$ are the conditional probabilities $P^*_{\mathrm{OFF}|n}$ and $P^*_{\mathrm{ON}|n}$, respectively. Substituting in (2.24) or (2.25) we get the recursive relation:

$$\frac{\mu_n(\alpha + c)}{\alpha + c + \beta\mu_n}P^*_{n+1} = \frac{\alpha}{\beta}P^*_n.$$

Finally, iterating and using the definitions of z and Z_c we find the probability mass (2.18).

The normalisation condition $\sum_n P^*_n = 1$ on the probability distribution (2.18) requires $\lim_{n\to\infty} \alpha/(\beta w_{c,n}) < 1$. For $\mu_n = \mu$, the effective departure rate is referred to as w_c, and this stationarity condition is simply $\alpha/(\beta w_c) < 1$. This implies that values of c smaller than the threshold

$$c_1 := \frac{\alpha^2}{\beta\mu - \alpha} \tag{2.27}$$

exclude any stationary state and produce a congested phase. The onset of congestion in a larger system with constant departure rate is explored in section 2.5 using a mean-field approach. For unbounded microscopic departure rates, i.e., $\lim_{n\to\infty} \mu_n = \infty$, the effective interaction is still bounded, since $\lim_{n\to\infty} w_{c,n} = (\alpha + c)/\beta$. However, as long as $c > 0$, the normalisation condition is always ensured, as $\lim_{n\to\infty} \alpha/(\beta w_{c,n}) = \alpha/(\alpha + c)$. Obviously, the linear departure rate case falls into this category.

2.4 The quantum Hamiltonian formalism

According to the section 1.2, the stochastic dynamics described by equations (2.3) and (2.4) can be represented by a Master equation in the form

$$\frac{\mathrm{d}}{\mathrm{d}t}|P(t)\rangle = -H|P(t)\rangle \tag{2.28}$$

with generator

$$-H = \begin{pmatrix} -\alpha-\delta & (\beta+\gamma)\mu_1 & 0 & 0 & \dots \\ \alpha+\delta & -\alpha-\delta-(\beta+\gamma)\mu_1 & -(\beta+\gamma)\mu_2 & 0 & \\ 0 & \alpha+\delta & -\alpha-\delta-(\beta+\gamma)\mu_2 & -(\beta+\gamma)\mu_3 & \\ 0 & 0 & \alpha+\delta & -\alpha-\delta-(\beta+\gamma)\mu_3 & \\ \vdots & & & & \ddots \end{pmatrix} \tag{2.29}$$

The choice of the letter H and the sign convention used are part of the so-called quantum Hamiltonian formalism, which is a convenient way to represent the dynamics of particle systems. The epithet "quantum" has become standard, along with the warnings that underline that the generator of a stochastic process is in general non-Hermitian, contrary to the operators in Quantum Mechanics. Also, at the risk of causing some confusion, the operator H is referred to as the Hamiltonian. For a classic account of such a formalism see reference [44]. The generic element $|e_n\rangle$ of the natural basis, which is a column vector with the n-th component equal to one and the remaining components equal to zero, represents a single-site configuration with n particles. Let define the creation and annihilation operators

$$a^+ = \begin{pmatrix} 0 & 0 & 0 & \dots \\ 1 & 0 & 0 & \\ 0 & 1 & 0 & \\ \vdots & & & \ddots \end{pmatrix} \quad \text{and} \quad a^- = \begin{pmatrix} 0 & \mu_1 & 0 & \dots \\ 0 & 0 & \mu_2 & \\ 0 & 0 & 0 & \\ \vdots & & & \ddots \end{pmatrix}, \tag{2.30}$$

respectively, and the diagonal operator

$$d = \begin{pmatrix} 0 & 0 & 0 & \dots \\ 0 & \mu_1 & 0 & \\ 0 & 0 & \mu_2 & \\ \vdots & & & \ddots \end{pmatrix}. \tag{2.31}$$

Now the Hamiltonian in equation (2.29) can be written as

$$H = -\alpha(a^+ - \mathbb{1}) - \delta(a^+ - \mathbb{1}) - \beta(a^- - d) - \gamma(a^- - d) \tag{2.32}$$

where the diagonal operators $\mathbb{1}$ and d encode for the conservation of the probability. Obviously, applying the ladder operator to the basis vectors yields:

$$a^+|e_n\rangle = |e_{n+1}\rangle \qquad \text{and} \qquad a^-|e_n\rangle = \mu_{n-1}|e_{n-1}\rangle. \tag{2.33}$$

These operators represent attempts rather than actual events; consistently the action of a^- on the empty site gives zero, i.e., there is no change in $|P(t)\rangle$. Let us now consider the ZRP with open boundaries on the one-dimensional lattice of length L, and represent the state as the tensor product

$$|P(t)\rangle = |P(t)\rangle_1 \otimes |P(t)\rangle_2 \otimes \ldots \otimes |P(t)\rangle_L, \tag{2.34}$$

where $|P(t)\rangle_l$ is the marginal probability distribution of the occupation of the l-th site, $l = 1, 2, \ldots, L$. We also define the ladder operators for each site, so that a_l^+, a_l^-, $\mathbb{1}_l$, and d_l are defined consistently with equations (2.30) and (2.31) but act non trivially on the l-th site, leaving the others unchanged. The jump of a particle from site l to site m is equivalent to the annihilation of a particle on site l and, simultaneously, the creation of a particle on site m. The total action can be written as the product operator $a_m^+ a_l^-$, which only modifies the m-th and l-th sites if the l-th site is non-empty, while gives zero otherwise. The boundaries are treated as in the single-site case. Consequently the complete Hamiltonian can be written compactly according to

$$\begin{aligned} -H = {} & \alpha(a_1^+ - \mathbb{1}_1) + \gamma(a_1^- - d_1) \\ & + \sum_{l=2}^{L-1} \left[p(a_l^+ a_{l-1}^- - d_{l-1}) + q(a_{l-1}^+ a_l^- - d_l)\right] + \beta(a_L^- - d_L) + \delta(a_L^+ - \mathbb{1}_L) \end{aligned} \tag{2.35}$$

We also mention that the formal solution of equation (2.28) can be expressed as

$$|P(t)\rangle = \mathrm{e}^{-Ht}|P(0)\rangle. \tag{2.36}$$

We now present the quantum Hamiltonian formalism for the on-off ZRP on the single site. We work in the joint occupation and phase vector space, defining a probability basis vector $|e_{n,\tau}\rangle = |e_n\rangle \otimes |e_\tau\rangle$ representing a configuration with n particles and phase τ. Hence, the generic probability vector can be written as $|P(t)\rangle = \sum_{n,\tau} P_{n,\tau}(t)|e_{n,\tau}\rangle$ Our convention is to use the basis kets $(1,0)^T$ and $(0,1)^T$ for the states $|e_{\mathrm{OFF}}\rangle$ and $|e_{\mathrm{ON}}\rangle$, respectively, while a configuration with n particles is represented by a basis ket with the n-th component equal to 1 and the remaining components equal to zero. Consequently, the Hamiltonian is written as

$$H = -c(a_{T_1}^+ - g_{T_1}) - \alpha(a_{N_1}^+ f_{T_1} - \mathbb{1}) - \beta(a_{N_1}^- d_{T_1} - d_{N_1} d_{T_1}), \tag{2.37}$$

with

$$\begin{aligned} a^+_{T_1} &= \mathbb{1} \otimes \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, f_{T_1} = \mathbb{1} \otimes \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \\ g_{T_1} &= \mathbb{1} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, d_{T_1} = \mathbb{1} \otimes \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \end{aligned} \tag{2.38}$$

$$a^+_{N_1} = \begin{pmatrix} 0 & 0 & 0 & \dots \\ 1 & 0 & 0 & \\ 0 & 1 & 0 & \\ \vdots & & & \ddots \end{pmatrix} \otimes \mathbb{1}, \tag{2.39}$$

$$a^-_{N_1} = \begin{pmatrix} 0 & \mu_1 & 0 & 0 & \dots \\ 0 & 0 & \mu_2 & 0 & \\ 0 & 0 & 0 & \mu_3 & \\ 0 & 0 & 0 & 0 & \\ \vdots & & & & \ddots \end{pmatrix} \otimes \mathbb{1}, \tag{2.40}$$

$$d_{N_1} = \begin{pmatrix} 0 & 0 & 0 & 0 & \dots \\ 0 & \mu_1 & 0 & 0 & \\ 0 & 0 & \mu_2 & 0 & \\ 0 & 0 & 0 & \mu_3 & \\ \vdots & & & & \ddots \end{pmatrix} \otimes \mathbb{1}. \tag{2.41}$$

We used the convention that a ladder operator with subscript N_1 or T_1 acts non-trivially only on the occupation or phase subspace respectively. The additional subscripts 1 underline that this Hamiltonian generates the dynamics for the single-site case. The resulting operator H in the joint space has block tridiagonal structure:

$$H = \left(\begin{array}{cc|cc|cc|c} c+\alpha & 0 & 0 & 0 & 0 & 0 & \dots \\ -c & \alpha & 0 & -\beta\mu_1 & 0 & 0 & \\ \hline -\alpha & -\alpha & c+\alpha & 0 & 0 & 0 & \\ 0 & 0 & -c & \alpha+\beta\mu_1 & 0 & -\beta\mu_2 & \dots \\ \hline \vdots & & & & & \vdots & \ddots \end{array}\right). \tag{2.42}$$

This type of representation is typical, in general, of stochastic generators of processes with two variables, such as n and τ in this case. The blocks correspond to changes in the first variable, while the entries inside the blocks correspond to changes of the second one. All the variables can change by at most 1. Such processes belong to the class of *quasi-birth–death* processes and are simple cases of queues with Markovian arrival and

Phase-type departure policy [42, 47]. We mention that the results in this section can be adapted to the more general PA case with the replacement $\alpha \to \alpha + \delta$ and $\beta \to \beta + \gamma$. Specifically, the quantum Hamiltonian for PA dynamics on a single site is

$$H = -c(a^+_{T_1} - g_{T_1}) - \alpha(a^+_{N_1} f_{T_1} - \mathbb{1}) - \beta(a^-_{N_1} d_{T_1} - d_{N_1} d_{T_1}) \\ - \gamma(a^-_{N_1} d_{T_1} - d_{N_1} d_{T_1}) - \delta(a^+_{N_1} f_{T_1} - \mathbb{1}). \quad (2.43)$$

On a many-site one-dimensional lattice of length L with open boundaries, the quantum Hamiltonian reads

$$H = -\sum_{l=1}^{L} c(a^+_{T_l} - g_{T_l}) - \sum_{l=1}^{L-1} \left[p(a^+_{N_{l+1}} f_{T_{l+1}} a^-_{N_l} d_{T_l} - d_{N_l} d_{T_l}) + q(a^+_{N_l} f_{T_l} a^-_{N_{l+1}} d_{T_{l+1}} - d_{N_{l+1}} d_{T_{l+1}})\right] + \\ - \alpha(a^+_{N_1} f_{T_1} - \mathbb{1}) - \beta(a^-_{N_L} d_{T_L} - d_{N_L} d_{T_L}) - \gamma(a^-_{N_l} d_{T_l} - d_{N_1} d_{T_1}) - \delta(a^+_{N_L} f_{T_L} - \mathbb{1}), \quad (2.44)$$

where the ladder operators $a^+_{N_l,T_l}$, $a^-_{N_l,T_l}$, d_{T_l}, g_{T_l}, and f_{T_l} have the same definitions as in equations (2.38)–(2.41) after the replacement $1 \to l$. The additional index l, where $l = 1, 2, \ldots, L$, indicates that the operator acts on the configuration n_l, τ_l of the site of index l. In the next section we will focus on an approximate stationary solution of the dynamics generated by this Hamiltonian.

2.5 Mean-field solution for the many-site system

In the case considered so far, particles arrive on the site from the boundaries according to a Poisson process. The many-site system is rather more complicated than this. In fact, each site receives particles according to a more general point process, which alternates time intervals with no events (corresponding to the off phases of the neighbour sites) and periods with arrivals. Moreover, the exact statistics of the phase switching is not *a priori* known.

In this section, we propose an approximate solution for the stationary state of the on-off ZRP on the one-dimensional lattice of length L with open boundaries. The approximation consists in decoupling the equations which describe the dynamics for each site, replacing the point process that governs the arrival on each site by a Poisson process with an effective characteristic rate. The decoupling of the equations allows us to use the results obtained for the one-site system in section 2.3.

Let us first consider the general model described in section 2.2, where the departure

rates $\mu_{n,\tau}$ retain a non trivial dependence on both n and τ. We are interested to write the exact Master equation for the probability $P_{n_l,\tau_l;l}(t)$ of having a configuration (n_l,τ_l) of the bulk site l. However, we cannot isolate the behaviour of site l from the remaining $L-1$ sites and it is more convenient to consider the joint probability $P_{\{n_j,\tau_j\}_j}(t)$ of having a configuration

$$\{n_j,\tau_j\}_j := ((n_1,\tau_1)_1,\ (n_2,\tau_2)_2,\ \ \ldots,\ (n_L,\tau_L)_L)\,, \tag{2.45}$$

where the j-th lattice site has occupation number n_j and phase τ_j. To make explicit the state of the site l, we also reserve the use of

$$\{n_j,\tau_j\}_{j\neq l},(n_l,\tau_l)_l \tag{2.46}$$

for the same configuration as (2.45). Summing the joint probability $P_{\{n_j,\tau_j\}_j}(t)$ over the occupation and clock states of all the sites except l, we obtain the marginal probability of having the l-th site in configuration (n_l,τ_l). With a slight abuse of notation this can be written as

$$P_{n_l,\tau_l;l}(t) = \left(\prod_{i\neq l}\cdot\sum_{n_i,\tau_i}\right) P_{\{n_j,\tau_j\}_j}(t). \tag{2.47}$$

In a similar fashion, we use $P_{n_l+1,\tau_l;l}(t)$ for the probability of having the same site l with n_l+1 particles and phase τ_l. We finally write:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}P_{n_l,\tau_l;l}(t) = {}& cP_{n_l,\tau_l-1;l}(t) + (p+q)\mu_{n_l+1,\tau_l}P_{n_l+1,\tau_l;l}(t) \\
&+ p\left(\prod_{i\neq l}\cdot\sum_{n_i,\tau_i}\right)\sum_{\tau}\mu_{n_{l-1},\tau}P_{\{n_j,\tau_j\}_{j\neq l},(n_l-1,\tau)_l}(t)\,\delta_{\tau_l,0} \\
&+ q\left(\prod_{i\neq l}\cdot\sum_{n_i,\tau_i}\right)\sum_{\tau}\mu_{n_{l+1},\tau}P_{\{n_j,\tau_j\}_{j\neq l},(n_l-1,\tau)_l}(t)\,\delta_{\tau_l,0} \\
-(p+q)\mu_{n_l,\tau_l}P_{n_l,\tau_l;l} - p\left(\prod_{i\neq l}\cdot\sum_{n_i,\tau_i}\right)&\mu_{n_l,\tau_l}P_{\{n_j,\tau_j\}_j}(t) - q\left(\prod_{i\neq l}\cdot\sum_{n_i,\tau_i}\right)\mu_{n_l,\tau_l}P_{\{n_j,\tau_j\}_j}(t).
\end{aligned} \tag{2.48}$$

We now assume a product measure $\prod_{i=1}^{L}P^*_{n_i,\tau_i}$ for the stationary joint probability $P^*_{\{n_j,\tau_j\}_j}$ and impose this solution in the stationarity condition $\mathrm{d}P_{n_l,\tau_l;l}(t)/\mathrm{d}t = 0$ for equa-

tion (2.48). This allows us to cancel the factor $\prod_{i\neq l} P^*_{n_i,\tau_i}$ to obtain

$$cP^*_{n_l,\tau_l-1;l} + p\sum_\tau z_{l-1}P^*_{n_l-1,\tau;l}\delta_{\tau_l,0} + q\sum_\tau z_{l+1}P^*_{n_l-1,\tau;l}\delta_{\tau_l,0} + (p+q)\mu_{n_l+1,\tau_l}P^*_{n_l+1,\tau_l;l} \\ = [pz_{l-1} + qz_{l+1} + (p+q)\mu_{n_l,\tau_l} + c]\, P^*_{n_l,\tau_l;l}, \quad (2.49)$$

for the generic bulk site $1 < l < L$. We use the symbol z_l, already adopted in section 2.3 for the fugacity, to denote the ensemble average of the departure rate, since $z_l = \sum_{n_l,\tau_l} \mu_{n_l,\tau_l} P^*_{n_l,\tau_l;l}$. This equivalence, in fact, is easily verified in the one-site system by means of equations (2.18)–(2.20) and remains consistent on chain topology. The use of an average interaction term justifies the appellation mean-field. Similarly, for the leftmost and rightmost sites we get, respectively,

$$cP^*_{n_1,\tau_1-1;1} + \alpha\sum_\tau P^*_{n_1-1,\tau;1}\delta_{\tau_1,0} + q\sum_\tau z_2 P^*_{n_1-1,\tau;1}\delta_{\tau_1,0} + (p+\gamma)\mu_{n_1+1,\tau_1}P^*_{n_1+1,\tau_1;1} \\ = [\alpha + qz_2 + (p+\gamma)\mu_{n_1,\tau_1} + c]P^*_{n_1,\tau_1;1} \quad (2.50)$$

and

$$cP^*_{n_L,\tau_L-1;L} + p\sum_\tau z_{L-1}P^*_{n_L-1,\tau;L}\delta_{\tau_L,0} + \delta\sum_\tau P^*_{n_L-1,\tau;L}\delta_{\tau_L,0} + (\beta+q)\mu_{n_L+1,\tau_L}P^*_{n_L+1,\tau_L;L} \\ = [pz_{L-1} + \delta + (\beta+q)\mu_{n_L,\tau_L} + c]P^*_{n_L,\tau_L;L}. \quad (2.51)$$

In equation (2.49) we recognise the stationarity condition for the single site with arrival and departure rate equal respectively to $(pz_{i-1} + qz_{i+1})$ and $(p+q)\mu_{n_i,\tau_i}$. Similarly, equation (2.50) is the stationarity condition for a single site with arrival and departure rates equal respectively to $(\alpha+qz_2)$ and $(p+\gamma)\mu_{n_1,\tau_1}$, while equation (2.51) has arrival and departure rates equal respectively to $(pz_{L-1}+\delta)$ and $(\beta+q)\mu_{n_L,\tau_L}$. Analogous equations hold when $\tau = 0$ and $n = 0$. These conditions, in the simplified case $\mu_{n,\tau} = \mu_n$ when $\tau > 0$, allow us to write an approximate stationary distribution for each site l analogous to (2.18) but with modified hopping rates

$$P^*_{n;l} = \frac{{z_l}^n}{Z_{c,l}} \prod_{i=1}^{n} w^{-1}_{c,i;l} \quad (2.52)$$

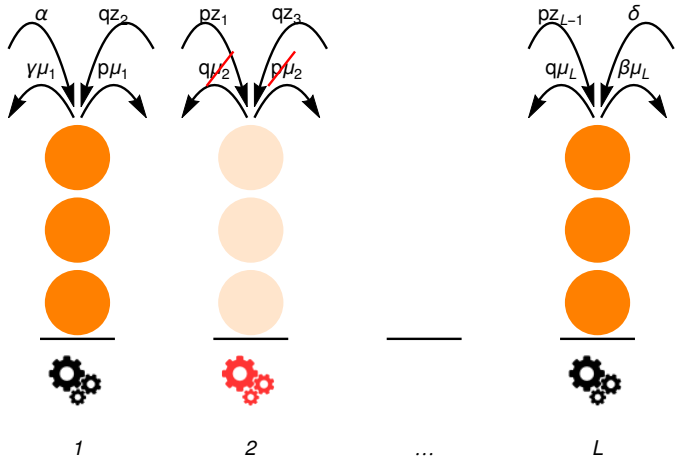

Figure 2.7: Representation of the on-off ZRP with open boundaries in mean-field approximation. Each lattice site is thought of as receiving particles according to the fugacity of its neighbours (which is constant in time), while on-site particles depart according to the true on-off mechanism.

with $Z_{c;l} = \sum_n {z_l}^n \prod_{i=1}^n w_{c,i;l}^{-1}$ and

$$\begin{aligned} z_1 &= \frac{\alpha + qz_2}{p+\gamma}, & w_{c,i;1} &= \frac{\mu_i(c+\alpha+qz_2)}{c+\alpha+qz_2+(p+\gamma)\mu_i}, \\ z_l &= \frac{pz_{l-1}+qz_{l+1}}{p+q}, & w_{c,i;l} &= \frac{\mu_i(c+pz_{l-1}+qz_{l+1})}{c+pz_{l-1}+qz_{l+1}+(p+q)\mu_i}, \\ z_L &= \frac{\delta + pz_{L-1}}{\beta+q}, & w_{c,i;L} &= \frac{\mu_i(c+\delta+pz_{L-1})}{c+\delta+pz_{L-1}+(\beta+q)\mu_i}, \end{aligned} \tag{2.53}$$

where $1 < l < L$. For constant departure rates, we can safely drop the label i from μ_i, $w_{c,i;1}$, $w_{c,i;l}$, and $w_{c,i;L}$. The scenario of equations (2.53) corresponds to a one-dimensional lattice where each site l receives a uniform (in time) particle stream of rate qz_{l+1} (pz_{l-1}) from the right (left) neighbour and sends particles according to its internal dynamics (with the condition that the total current along the chain is conserved), as illustrated in figure 2.7. In the quantum Hamiltonian formalism, the approximation results in the separation of the two-site bulk terms of equation (2.44), i.e., $p(a^+_{N_{l+1}} f_{T_{l+1}} a^-_{N_l} d_{T_l}) \sim pz_l(a^+_{N_{l+1}} f_{T_{l+1}} + a^-_{N_l} d_{T_l})$, and $q(a^+_{N_l} f_{T_l} a^-_{N_{l+1}} d_{T_{l+1}}) \sim qz_{l+1}(a^+_{N_l} f_{T_l} + a^-_{N_{l+1}} d_{T_{l+1}})$. The approximation results in the separation of the dynamics of each site, consequently, the mean-field quantum Hamiltonian can be written as:

$$H_{\text{mf}} = H_{\text{left}} + H_{\text{right}} + \sum_{l=2}^{L-1} H_l, \tag{2.54}$$

where H_{left}, H_{right}, and H_l are obtained from the generic one-site Hamiltonian (2.43) using the mean-field arrival rates.

The fugacities z_l can be calculated self-consistently from their definition in equations (2.53). However, to find their values, it is more convenient to impose the conservation of the current along the chain, as has already been done for the standard ZRP on a chain [33] in terms of the quantum Hamiltonian formalism. The first observation is that the stationary state satisfies the following relations:

$$\begin{aligned} a_l^+|P^*\rangle_l &= z_l^{-1} d_l |P^*\rangle_l \\ a_l^-|P^*\rangle_l &= z_l |P^*\rangle_l, \end{aligned} \tag{2.55}$$

where is $|P^*\rangle_l$ the probability distribution of the occupation of the l-th site, $i = 1, 2, \ldots, L$. Substituting back into the Master equation, we obtain the stationarity condition

$$\begin{aligned} -H|P^*\rangle = (\alpha+qz_2)(z_1^{-1}d_1-1)+(\gamma+p)(z_1-d_1)+(pz_1+qz_3)(z_2^{-1}d_2-1)+(p+q)(z_2-d_2) \\ + \ldots + (pz_{L-1}+\delta)(z_L^{-1}d_L - 1) + (\beta+q)(z_L - d_L)|P^*\rangle = 0, \end{aligned} \tag{2.56}$$

which is satisfied for

$$\alpha - \gamma z_1 = p z_l - q z_{l+1} = \beta z_L - \delta = \langle j \rangle. \tag{2.57}$$

$l = 1, 2, \ldots, L$. It is proved in reference [33] that the solution of this recursive relation yields the fugacity z_l and the mean current $\langle j \rangle$:

$$z_l = \frac{\alpha\beta\left(\frac{p}{q}\right)^{L-1} - \gamma\delta - \left(\frac{p}{q}\right)^{l-1}[\alpha\beta - \gamma\delta - (\alpha+\delta)(p-q)]}{\beta\left(\frac{p}{q}\right)^{L-1}(\gamma+p-q) - \gamma(\beta-p+q)}, \tag{2.58}$$

$$\langle j \rangle = (p-q)\frac{\alpha\beta\left(\frac{p}{q}\right)^{L-1} - \gamma\delta}{\gamma(p-q-\beta) + \beta(p-q+\gamma)\left(\frac{p}{q}\right)^{L-1}}, \tag{2.59}$$

which complete the mean field solution for the model.

While the fugacities (2.58) are identical to those of a standard ZRP on an open chain [33], the effective departure rates $w_{c,n;l}$ are affected by the time correlations (they retain a dependence on the parameter c) and, significantly, become site dependent. This is evident at the level of stationary density and variance profile, respectively $\langle n_l \rangle = z_l \partial(\ln Z_{c,l})/\partial z_l$ and $\sigma_l{}^2 \equiv \langle {n_l}^2 \rangle - \langle n_l \rangle^2 = {z_l}^2 \partial^2(\ln Z_{c,l})/\partial {z_l}^2$ in the mean-field approximation. The predicted density profile can be non-monotonic, which contrasts with the stationary profile of the standard ZRP [33]. This feature is indeed present in the

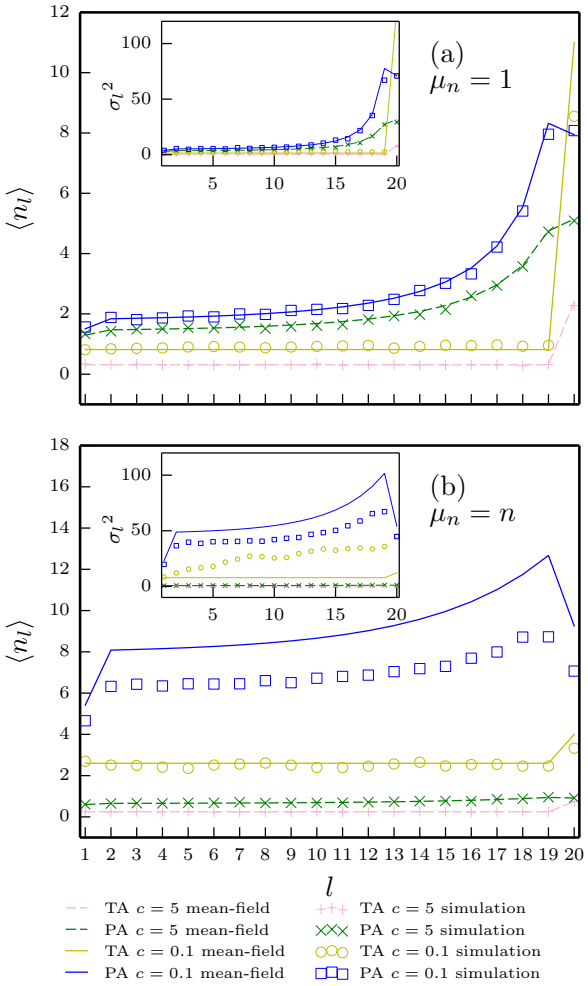

Figure 2.8: Density profile $\langle n_l \rangle$, and variance profile ${\sigma_l}^2$ (insets), $l = 1, 2, \ldots, L$ for the on-off dynamics on a chain of length $L = 20$ with (a) constant and (b) linear departure rates. The results obtained within the mean-field approximation (line vertices) are compared with the simulation results obtained by means of the Doob–Gillespie method of section 1.7 (markers). TA and PA refer to rates $(\alpha, \beta, \gamma, \delta, p, q) = (0.2, 0.3, 0, 0, 1, 0)$ and $(\alpha, \beta, \gamma, \delta, p, q) = (0.1, 0.2, 0.1, 0.1, 0.55, 0.45)$ respectively.

Monte Carlo simulated density profiles for certain parameter combinations. In fact, for the parameters considered, the agreement between mean-field theory and simulation is excellent, except when c is very small, as shown in figure 2.8. For all the sets of parameters considered, the agreement increases with c.

In figure 2.9 the mean-field predictions for the per-site occupation distributions are also checked against Monte Carlo simulations. Both mean-field and simulation results show an exponential decay of the tails which decreases as the site index l increases. The agreement is again good, except for the cases with smaller values of c. In fact, in contrast to the mean-field prediction (2.52), the empirical distributions are glaringly non-Geometric when the parameter c is small enough.

The failure of the product solution (2.1) is also reflected in the behaviour of the cross-correlation $C_{ij} = (\langle n_i n_j \rangle - \langle n_i \rangle \langle n_j \rangle)/(\sigma_i \sigma_j)$ between the occupations on site i and j. This is shown in figure 2.10, where we report a negative cross-correlation between

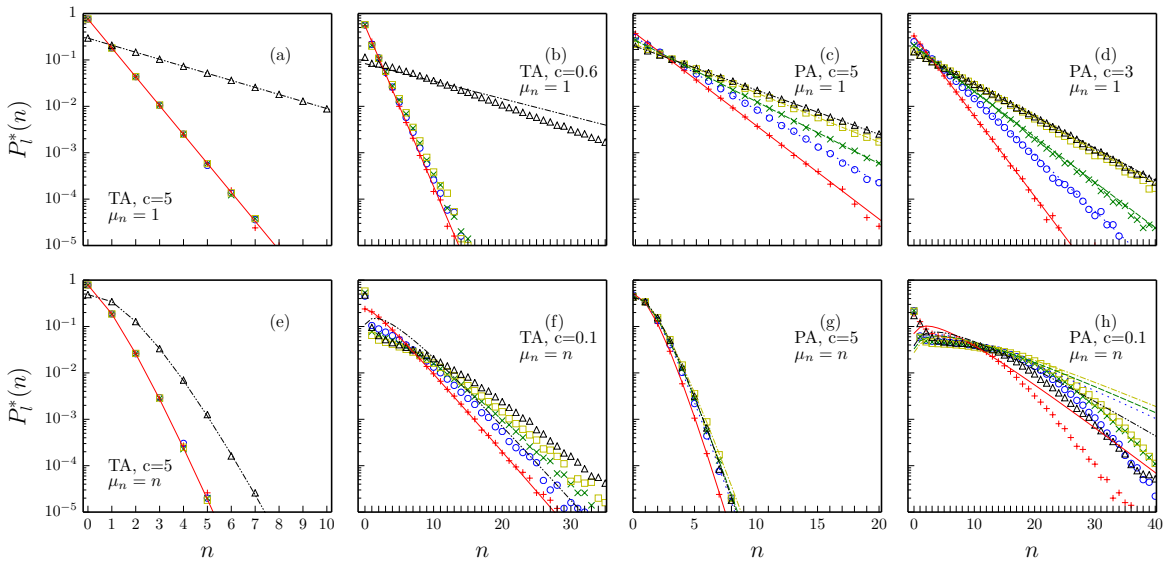

Figure 2.9: The mean-field probability distribution (line vertices) of the site occupation numbers in a chain of length $L = 5$ are checked against simulations (markers). Symbols $+, \times, \circ, \square, \triangle$ and solid, dotted, dashed, dot–dashed, dot–dot–dashed lines of the corresponding color (grayscale) refer to sites $l = 1, 2, 3, 4, 5$ respectively. Parameter combinations as in figure 2.8.

neighbouring-site occupations for small values of c.

2.6 Congestion threshold

In the cases explored above, the values of c have been chosen in order to guarantee the existence of a well defined NESS with constant average occupation number. In the one-site system seen in section 2.3 this choice was straightforward, as we can derive exactly the congestion threshold. In an extended system with unbounded departure rates, we expect that any strictly positive value of c guarantees the NESS because, although a large number of particles can pile up during the off phase, they can be released arbitrarily quickly during the on phase. On the contrary, the extended system with bounded departure rates appears to be more interesting. In fact, for values of c smaller than a certain value, the particles accumulate on one or more of the lattice sites. We now compare the prediction of the mean-field theory for this congestion threshold with the results of Monte Carlo simulations performed on a chain of length $L = 20$. In order to evaluate numerically the onset of congestion, we make use of the parameter

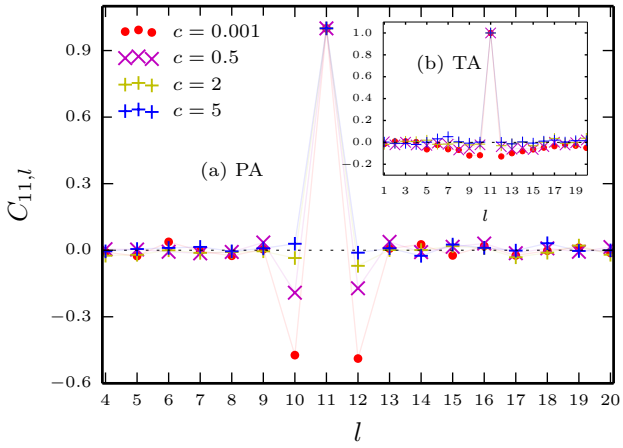

Figure 2.10: Simulation results for the cross-correlation $C_{11,l}$ of on-off ZRP on a chain of length $L = 20$ with $\mu_n = n$. (a) PA case. Adjacent sites have negatively-correlated occupation numbers. Hopping-rate combinations as in figure 2.8. Inset, (b). TA case. Spatial cross-correlations appear weaker than the PA case, but with longer range. As c grows the correlation is gradually lost and a factorised solution is realistic.

(inspired by reference [3])

$$\kappa = \frac{n_{\text{tot}}(t + \Delta t) - n_{\text{tot}}(t)}{\Delta t} \frac{1}{(\alpha + \delta)}, \tag{2.60}$$

where $t \gg \Delta t$, $n_{\text{tot}}(t)$ is the average total number of particles in the system at time t. The parameter κ measures the difference between the rate at which particles arrive and the rate at which particles leave the system, scaled with respect to the total arrival rate. The congestion occurs when κ is strictly positive[2]. For the one-site model (2.14)–(2.17) with $\mu_n = \mu$, it is straightforward to show that the value of $\alpha\kappa$ is

$$\alpha\kappa = \begin{cases} \alpha - \mu\beta c/(c + \alpha) & \text{if } c < c_1, \\ 0 & \text{elsewhere,} \end{cases} \tag{2.61}$$

where c_1 is given by equation (2.27). In fact, in the congested regime we can assume that the site is never empty and the total growth rate is given by the difference $\alpha - \mu\beta P^*_{\text{ON}}$, where $P^*_{\text{ON}} = \sum_{n=0}^{\infty} P^*_{n,\text{ON}}$ is the stationary probability of having the site in the state on. Summing over n the equations (2.14) and (2.16), and imposing the stationarity conditions, we get $P^*_{\text{ON}} = c/(c + \alpha)$.

Equation (2.61) allows us to approximate a local κ_l for the generic site l of a chain,

[2]Precisely at the threshold, we expect congestion/condensation but with sublinear growth in time.

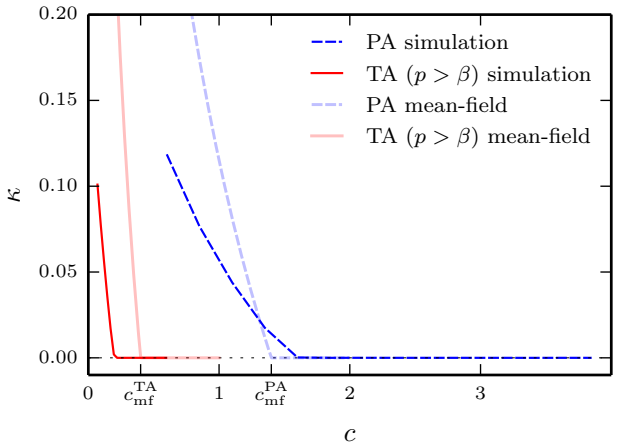

Figure 2.11: Congestion transition for the on-off ZRP with $\mu = 1$ on a open chain of length $L = 20$. The parameter κ obtained by Monte Carlo simulations is plotted against c for a TA case and a PA case. Hopping-rate combinations as in figure 2.8. The mean-field congestion thresholds are $c_{\rm mf}^{\rm TA} = 0.4$ and $c_{\rm mf}^{\rm PA} \simeq 1.4$ respectively. These values are pinpointed by the mean-field approximated local κ_l (light lines) of the site l where the congestion sets in first.

by replacing α and β with the mean-field arrival and departure rates, respectively. The numerical Monte Carlo study of κ_l reveals the first site where the congestion sets in, tuning c from large to smaller values. In a chain with TA jumps, this seems to occur on site 1, for $p \leq \beta$, or on the site L, otherwise. In the PA case, the congestion can set in on the bulk site $L - 1$, as suggested by the non-monotonic density profile of figure 2.8. Not surprisingly, the mean-field theory predicts this possibility.

We define the mean-field congestion threshold $c_{\rm mf}$ as the smallest value of c such that none of the sites l of the system with Hamiltonian (2.54) has $\kappa_l > 0$. In the TA case, as long as $p > \beta$, $c_{\rm mf}$ is equivalent to the threshold c_1 of equation (2.27), derived for the one-site system with boundary rates α and β. The numerical evaluation of κ for the whole system, plotted against the mean-field estimate in figure 2.11, suggests that $c_{\rm mf}$ is an upper bound for the true congestion transition in this case. We argue that this relation arises as the TA jumps set the system in a highly organised configuration, with wave-like fronts which are precursors of the slinky motion observed in Hirschberg et al. [28, 29] and enhance the particle transport. This fact can be deployed, to some extent, to design real-life queuing systems. When $p < \beta$, $c_{\rm mf} = \alpha^2/(p\mu - \alpha)$ marks exactly the onset of the congested phase. Conversely, PA interactions seem to promote congestion, as there are many jumps which block the site and contribute negatively to the particle current. In this case the congestion transition occurs for a value of c larger than both c_1

and c_{mf} (see figure 2.11).

2.7 Discussion

In this chapter, we commenced the investigation of the effects of temporal correlations on stochastic systems. Specifically, we focused on a variant of the ZRP, namely the on-off ZRP, that incorporates memory by means of hidden variables. In the on-off ZRP, the hidden variables affect the on-site dynamics by blocking the particle jumps for a certain amount of time which is tuned by the parameter c: the smaller its value is, the stronger the memory effects of the hidden dynamics are. In the limit as $c \to \infty$, the standard ZRP is recovered.

In the single-site case, we found the exact NESS, which can be written in terms of a standard ZRP process, with effective departure rate. Moving to the many-site case, one of the major consequences of the on-off interaction is the lack of a simple factorised form (as in equation (2.1)) for the stationary state, while it exists for the standard ZRP. As an approximation, we derived a mean-field stationary-state solution where the statistics of the particle jumps from site to site in a lattice are smoothed over time, while preserving the lattice topology. Such an approximation restores the product-form solution and makes excellent predictions when the value of c is large enough.

The mean-field approach gives also an analytical approximation for the value of c at which the system enters a congested phase. This prediction is particularly accurate for the case with totally symmetric interactions. On the contrary, in the partially asymmetric system, the mean-field congestion threshold underestimates the result suggested by simulations. The particle density and variance profiles within this approximations do not fit the results from the simulation for values of c very close to c_{mf}. However, for larger values of c (but still of the same order of c_{mf}), the mean-field predictions are in excellent agreement with the simulations.

This section also introduced interesting analogies between IPSs and Queuing networks. Due to their common ground, our results could be exploited by the Queueing Theory community and could lead to interesting applications, where on and off phases can be used to improve packet traffic.

The main message of this chapter can be summarised as follows. For small values of c the behaviour of NESS is heavily altered; however, the value of the mean current $\langle j \rangle$ across the chain only depends on the jump rates for a wide range of parameters, regardless of the value of c (as long as the congestion does not set in). This motivates

the study of rare fluctuations of the current, which is the topic of the next chapter.

Bibliography

[1] T. Antal, P. L. Krapivsky, and S. Redner. Shepherd model for knot-limited polymer ejection from a capsid. *J. Theo. Biol.*, 261(3):488–93, 2009.

[2] A. Arazi, E. Ben-Jacob, and U. Yechiali. Bridging genetic networks and queueing theory. *Physica A*, 332:585–616, 2004.

[3] A. Arenas, A. Díaz-Guilera, and R. Guimerà. Communication in networks with hierarchical branching. *Phys. Rev. Lett.*, 86(14):3196–3199, 2001.

[4] F. Baskett, K. M. Chandy, R. R. Muntz, and F. G. Palacios. Open, closed, and mixed networks of queues with different classes of customers. *J. ACM*, 22(2):248–260, 1975.

[5] G. Bianconi and A. Barabási. Bose-Einstein condensation in complex networks. *Phys. Rev. Lett.*, 86(24):5632–5635, 2001.

[6] R. A. Blythe and M. R. Evans. Nonequilibrium steady states of matrix-product form: a solver's guide. *J. Phys. A: Math. Gen.*, 40(46):R333–R441, 2007.

[7] J. Bouchaud and M. Mézard. Wealth condensation in a simple model of economy. *Physica A*, 282(3-4):536–545, 2000.

[8] Z. Burda, D. Johnston, J. Jurkiewicz, M. Kamiński, M. A. Nowak, G. Papp, and I. Zahed. Wealth condensation in pareto macroeconomies. *Phys. Rev. E*, 65(2): 026102, 2002.

[9] M. Cavallaro, R. J. Mondragón, and R. J. Harris. Temporally correlated zero-range process with open boundaries: Steady state and fluctuations. *Phys. Rev. E*, 92(2): 022137, 2015.

[10] V. Y. Chernyak, M. Chertkov, D. A. Goldberg, and K. Turitsyn. Non-equilibrium statistical physics of currents inqueuing networks. *J. Stat. Phys.*, 140(5):819–845, 2010.

[11] T. Chou and D. Lohse. Entropy-driven pumping in zeolites and biological channels. *Phys. Rev. Lett.*, 82(17):3552–3555, 1999.

[12] T. Chou, K. Mallick, and R. K. P. Zia. Non-equilibrium statistical mechanics: from a paradigmatic model to biological transport. *Rep. Progr. Phys.*, 74(11):116601, 2011.

[13] D. De Martino, L. Dall'Asta, G. Bianconi, and M. Marsili. Congestion phenomena on complex networks. *Phys. Rev. E*, 79(1):015101, 2009.

[14] D. De Martino, L. Dall'Asta, G. Bianconi, and M. Marsili. A minimal model for congestion phenomena on complex networks. *J. Stat. Mech.*, 2009(08):P08023, 2009.

[15] B. Derrida. An exactly soluble non-equilibrium system: The asymmetric simple exclusion process. *Phys. Rep.*, 301(1-3):65–83, 1998.

[16] K. R. Duffy and A. Sapozhnikov. The large deviation principle for the on-off Weibull sojourn process. *J. Appl. Probab.*, 45(1):107–117, 2008.

[17] J. Eggers. Sand as Maxwell's Demon. *Phys. Rev. Lett.*, 83(25):5322–5325, 1999.

[18] M. R. Evans and T. Hanney. Nonequilibrium statistical mechanics of the zero-range process and related models. *J. Phys. A*, 38(19):R195–R240, 2005.

[19] H. Fröhlich. Evidence for Bose condensation-like excitation of coherent modes in biological systems. *Phys. Lett. A*, 51(1):21–22, 1975.

[20] E. Gelenbe. Random neural networks with negative and positive signals and product form solution. *Neural Comput.*, 1(4):502–510, 1989.

[21] E. Gelenbe. Product-form queueing networks with negative and positive customers. *J. Appl. Prob.*, 28(3):656, 1991.

[22] E. Gelenbe. Learning in the recurrent random neural network. *Neural Comput.*, 5 (1):154–164, 1993.

[23] E. Gelenbe. G-networks: a unifying model for neural and queueing networks. *Ann. Oper. Res.*, 48(5):433–461, 1994.

[24] E. Gelenbe. Steady-state solution of probabilistic gene regulatory networks. *Phys. Rev. E*, 76(3):031903, 2007.

[25] E. Gelenbe and G. Pujolle. *Introduction to queueing networks*. Wiley, New York NY, 1998.

[26] O. Golinelli and K. Mallick. The asymmetric simple exclusion process: an integrable model for non-equilibrium statistical mechanics. *J. Phys. A: Math. Gen.*, 39(41): 12679–12705, 2006.

[27] S. Grosskinsky, F. Redig, and K. Vafayi. Dynamics of condensation in the symmetric inclusion process. *Elec. J. Prob.*, 18:1–23, 2013.

[28] O. Hirschberg, D. Mukamel, and G. M. Schütz. Condensation in temporally correlated zero-range dynamics. *Phys. Rev. Lett.*, 103(9):090602, 2009.

[29] O. Hirschberg, D. Mukamel, and G. M. Schütz. Motion of condensates in non-Markovian zero-range dynamics. *J. Stat. Phys.*, 2012(08):31, 2012.

[30] J. R. Jackson. Jobshop-like queueing systems. *Manag. Sci.*, 10(1):131–142, 1963.

[31] J. Kaupužs, R. Mahnke, and R. J. Harris. Zero-range model of traffic flow. *Phys. Rev. E*, 72(5):056125, 2005.

[32] S. Klein, C. Appert-Rolland, and L. Santen. Environmental control of microtubule-based bidirectional cargo transport. *EPL*, 107(1):18004, 2014.

[33] E. Levine, D. Mukamel, and G. M. Schütz. Zero-range process with open boundaries. *J. Stat. Phys.*, 120(5-6):759–778, 2005.

[34] T. M. Liggett. *Interacting Particle Systems*, volume 276 of *Grundlehren der mathematischen Wissenschaften*. Springer, New York NY, New York, NY, 1985.

[35] T. M. Liggett. Stochastic models of interacting systems. *The Annals of Probability*, (June 1996):1–29, 1997.

[36] T. M. Liggett. *Stochastic Interacting Systems: Contact, Voter and Exclusion Processes*, volume 324 of *Grundlehren der mathematischen Wissenschaften*. Springer,

Berlin Heidelberg Germany, 1999.

[37] T. M. Liggett. Stochastic models for large interacting systems and related correlation inequalities. *Proc. Nat. Ac. Sci.*, 107(38):16413–9, 2010.

[38] C. Maes. New trends in interacting particle systems. In *Markov Proc. Rel. Fields 11*, 2005.

[39] K. Mallick. The exclusion process: A paradigm for non-equilibrium behaviour. *Physica A*, 418:17–48, 2015.

[40] M. K. Mitra and S. Chatterjee. Boundary induced phase transition with stochastic entrance and exit. *J. Stat. Mech.*, 2014(10):P10019, 2014.

[41] R. J. Mondragón, D. K. Arrowsmith, and J. M. Pitts. Chaotic maps for traffic modelling and queueing performance analysis. *Perf. Eval.*, 43(4):223–240, 2001.

[42] M. F. Neuts. *Matrix-geometric Solutions in Stochastic Models: An Algorithmic Approach*. Dover Publications, New York NY, 1981.

[43] A. Schadschneider, D. Chowdhury, and K. Nishinari. *Stochastic Transport in Complex Systems: From Molecules to Vehicles*. Elsevier, Oxford UK, 2010.

[44] G. M. Schütz. Exactly solvable models for many-body systems far from equilibrium. *Phase transitions and critical phenomena*, 19:1–251, 2001.

[45] A. K. Sharma and D. Chowdhury. Stochastic theory of protein synthesis and polysome: Ribosome profile on a single mRNA transcript. *J. Theor. Biol.*, 289: 36 – 46, 2011.

[46] F. Spitzer. Interaction of Markov processes. *Adv. Math.*, 5(2):246–290, 1970.

[47] W. J. Stewart. *Probability, Markov Chains, Queues, and Simulation: The Mathematical Basis of Performance Modeling*. Princeton University Press, Princenton NJ, 2009.

[48] J. Török. Analytic study of clustering in shaken granular material using zero-range processes. *Physica A*, 355(2-4):374–382, 2005.

[49] H. Touchette, R. J. Harris, and J. Tailleur. First-order phase transitions from poles in asymptotic representations of partition functions. *Phys. Rev. E*, 81(3):030101, 2010.

[50] P. Trapman and M. C. J. Bootsma. A useful relationship between epidemiology and queueing theory: The distribution of the number of infectives at the moment of the first detection. *Math. Biosci.*, 219(1):15–22, 2009.

[51] A. M. J. VanDongen. K channel gating by an affinity-switching selectivity filter. *Proc. Nat. Acad. Sci. U.S.A.*, 101(9):3248–52, 2004.

[52] B. Waclaw and M. R. Evans. Explosive condensation in a mass transport model. *Phys. Rev. Lett.*, 108(7):070601, 2012.

[53] C. N. Yang and T. D. Lee. Statistical theory of equations of state and phase transitions. i. theory of condensation. *Phys. Rev.*, 87(3):404–409, 1952.

3 Temporally correlated zero-range process with open-boundary: fluctuations

Contents

This chapter is devoted to the study of particle-current fluctuations for the model introduced in chapter 2. Such fluctuations are quantified by means of the SCGF $e(s)$ as detailed in section 1.6. In section 3.1 we explore analytically and numerically the current fluctuations on a one-site system, focusing on the difference between the small fluctuation regime, obtained by analytic continuation of the NESS (subsection 3.1.1), and the extreme fluctuation regimes (subsection 3.1.3), and explaining the mechanisms that separate the phases (subsection 3.1.2).

Along the lines of the previous chapter, we focus on two simple functional forms for the particle-particle interaction factor, viz., linear and constant with the site occupation number n. This second case appears analytically more challenging and its SCGF is obtained as the asymptotic limit of the integral representation of an approximated moment generating function (developed throughout subsection 3.1.3). The analytical results are tested against the numerical method of Lecomte and Tailleur [12] (as described in section 1.7.2), which is also used to measure the bond-current fluctuations for the many-site model in section 3.2. Akin to the one-site system behaviour, in the many-site system it is

possible to separate a small-fluctuation regime, which is not influenced by the temporal correlations, from the extreme-current regimes where memory plays an important role.

3.1 Exact results for the one-site system

Despite its simplicity, the single-site ZRP exhibits a rich fluctuating behaviour, even in the absence of time correlations, as shown in references [5–7, 14]. The introduction of the on-off dynamics creates a still more interesting scenario. In fact, the study of the fluctuations reveals some aspects of the correlations which, in the stationary state, are hidden within an effective interaction factor. Throughout this chapter we assume that the observation starts at a time $t_0 = 0$, hence the length of the observation time is t. We denote by j the empirical output current J/t, i.e., the difference J between the number of jumps of particles which leave the site rightwards (with rate β) and the number of particle injections from the right boundary (which occur with rate δ) divided by the observation time. The input current, i.e., the rightwards current between the left boundary and the site, can be obtained by reflection in the PA case (while in the TA case this is given by a simple Poisson process).

The strategy outlined in section 1.6 is to build an s-modified generator, where the rates that contribute an increase or decrease in the current are multiplied by a factor e^{-s} or e^{s}, respectively. Our choice is to work with the Hamiltonian (2.43), where such rates are grouped into the ladder operators $\beta a^{-}_{N_1}$ and $\delta a^{+}_{N_1}$. Consequently, the s-modified Hamiltonian corresponding to the output current is

$$\begin{aligned}\tilde{H} = -c(a^{+}_{T_1} - g_{T_1}) - \alpha(a^{+}_{N_1} f_{T_1} - \mathbb{1}) - \beta(\mathrm{e}^{-s} a^{-}_{N_1} d_{T_1} - d_{N_1} d_{T_1}) \\ - \gamma(a^{-}_{N_1} d_{T_1} - d_{N_1} d_{T_1}) - \delta(\mathrm{e}^{s} a^{+}_{N_1} f_{T_1} - \mathbb{1}).\end{aligned} \tag{3.1}$$

We concentrate now on the eigenproblem

$$(\tilde{H} - A\mathbb{1})|\tilde{P}_A\rangle = 0, \tag{3.2}$$

where $|\tilde{P}_A\rangle$ is the generic right eigenvector and A is its eigenvalue. Specifically, we are first interested in the leading eigenvalue A_0, which is proven to be equal to the SCGF, at least in the neighbourhood of $s = 0$. Notice that the quantum Hamiltonian formalism uses a sign convention different than the one of section 1.6; in fact, its leading eigenvalue A_0 corresponds to the leading eigenvalue λ_1 of the generator $\tilde{\mathbf{G}}$ (but has opposite sign) and is the smallest eigenvalue of $\tilde{H}$ (at least in the neighbourhood of $s = 0$). The associated right eigenvector is also called ground-state eigenvector.

3.1.1 Small current fluctuations

The working hypothesis is that the eigenvector $|\tilde{P}_{A_0}\rangle$, associated to A_0, has a form similar to the stationary solution (2.18)–(2.20). To lighten the notation, we drop the subscript A_0 and use $|\tilde{P}\rangle$ and $\langle\tilde{P}|$ for the right and left eigenvectors associated to A_0, respectively. It is convenient to write the ground-state eigenvector $|\tilde{P}\rangle$ in a form similar to the stationary solution (2.18)–(2.20), i.e., with components

$$\tilde{P}_{n,\mathrm{ON}} = p_{\mathrm{ON},n,s}\tilde{P}_n, \tag{3.3}$$

$$\tilde{P}_{n,\mathrm{OFF}} = (1 - p_{\mathrm{ON},n,s})\tilde{P}_n, \tag{3.4}$$

$$\tilde{P}_{n+1} = \rho_{n+1,s}\tilde{P}_n. \tag{3.5}$$

Equation (3.2) is hard to solve in general. To gain insight into the appropriate structure of $\rho_{n,s}$ and $p_{\mathrm{ON},n,s}$, we study first the case with constant departure rates.

3.1.1.1 Constant departure rates

Let the departure rate be $\mu_n = \mu$ when $n > 0$ (obviously, there are no departures when $n = 0$, i.e., $\mu_0 = 0$). Motivated by the stationary-state result, we assume here that the factors $p_{\mathrm{ON},n,s}$ and $\rho_{n,s}$ have no dependence on the occupation number and we drop the subscript n with the exception of $n = 0$, i.e., $p_{\mathrm{ON},0,s}$ is distinct from $p_{\mathrm{ON},s}$. By direct substitution into equation (3.2) we get:

$$-(c + \alpha + \delta - A_0)(1 - p_{\mathrm{ON},0,s}) = 0, \tag{3.6}$$

$$c(1 - p_{\mathrm{ON},0,s}) - (\alpha + \delta - A_0)p_{\mathrm{ON},0,s} + (\beta \mathrm{e}^{-s} + \gamma)\mu p_{\mathrm{ON},s}\rho_s = 0, \tag{3.7}$$

$$(\alpha + \delta \mathrm{e}^{s}) - (c + \alpha + \delta - A_0)(1 - p_{\mathrm{ON},s})\rho_s = 0, \tag{3.8}$$

$$c(1 - p_{\mathrm{ON},s}) - [\alpha + \delta + (\beta + \gamma)\mu - A_0]p_{\mathrm{ON},s} + (\beta e^{-s} + \gamma)\mu p_{\mathrm{ON},s}\rho_s = 0. \tag{3.9}$$

Equation (3.6) trivially requires $p_{\mathrm{ON},s,0} = 1$, while we expect that $p_{\mathrm{ON},s} < 1$. After a long but straightforward algebraic manipulation, the system is solved as

$$p_{\mathrm{ON},s} = \frac{c + (\beta \mathrm{e}^{-s} + \gamma)(\alpha + \delta \mathrm{e}^{s})/(\beta + \gamma)}{c + (\beta + \gamma)\mu + (\beta \mathrm{e}^{-s} + \gamma)(\alpha + \delta \mathrm{e}^{s})/(\beta + \gamma)} \tag{3.10}$$

$$\rho_s = \frac{(\alpha + \delta \mathrm{e}^{s})}{(\beta + \gamma)}(\mu p_{\mathrm{ON},s})^{-1}, \tag{3.11}$$

$$A_0 = \frac{\alpha\beta}{\beta + \gamma}(1 - \mathrm{e}^{-s}) + \frac{\gamma\delta}{\beta + \gamma}(1 - \mathrm{e}^{s}). \tag{3.12}$$

Note that setting $s = 0$, the factor $p_{\mathrm{ON},s}$ becomes the conditional probability $P^*_{\mathrm{ON}|n}$ in

the steady state. Also, the parameter ρ_s and the eigenvalue A_0 have a counterpart in the stationary probability, in fact for $s \to 0$, $\rho_s \to z w_c^{-1}$, and $A_0 \to 0$. Consequently, we argue that A_0 is the lowest eigenvalue of $\tilde{H}$ and, according to section 1.6, the SCGF at least in the neighbourhood of $s = 0$.

For later convenience, we define a modified fugacity

$$z_s = \frac{\alpha + \delta \mathrm{e}^s}{\beta + \gamma} \tag{3.13}$$

and a modified effective interaction

$$w_{c,s} = \mu p_{\mathrm{ON},s}, \tag{3.14}$$

such that $\rho_s = z_s w_{c,s}^{-1}$ and $w_{c,s} \to \mu$ for $c \to \infty$. It is worth noting that, while the bias affects only the fugacity in the ordinary ZRP [5], it affects both the interaction term and the fugacity in the on-off model.

3.1.1.2 General departure rates

The solution (3.10) and (3.11) suggests an ansatz for the case with arbitrary dependence of the departure rates on n, which is now considered. This also covers the special case with linear departure rates $\mu_n = n$. Motivated by the results for constant $\mu_n = \mu$, we assume that the components of the ground-state eigenvector $|\tilde{P}\rangle$ satisfy equations (3.3)–(3.5) with

$$\rho_{n,s} = z_s w_{c,n,s}^{-1}, \tag{3.15}$$

$$w_{c,n,s} = \mu_n p_{\mathrm{ON},n,s}, \tag{3.16}$$

for $n \geq 0$. With this assumption, the second row equation of the eigenproblem (3.2) is solved for $A = A_0 \equiv \alpha + \delta - (\beta \mathrm{e}^{-s} + \gamma) z_s$ and the remaining equations yield a solution for z_s consistent with (3.13) and an n-dependent effective interaction

$$w_{c,n,s} = \frac{\mu_n [c + (\beta \mathrm{e}^{-s} + \gamma)(\alpha + \delta \mathrm{e}^s)/(\beta + \gamma)]}{(\beta \mathrm{e}^{-s} + \gamma)(\alpha + \delta \mathrm{e}^s)/(\beta + \gamma) + c + (\beta + \gamma)\mu_n}. \tag{3.17}$$

A comparison with the constant departure rate case shows that $w_{c,n,s}$ can be obtained from $w_{c,s}$, replacing μ with μ_n.

The eigenvalue we obtained is the same as the lowest eigenvalue A_0 (3.12) of the s-Hamiltonian for the standard ZRP [5]. In fact, the affinity between the two models

appears closer if we work in the reduced state space obtained by collapsing the states corresponding to $\tau = \text{ON}$ and $\tau = \text{OFF}$, for each occupation number, and considering the sum of their non-conserved probabilities $\tilde{P}_n = \tilde{P}_{n,\text{ON}} + \tilde{P}_{n,\text{OFF}}$. We represent the column vector of components $\tilde{P}_n$ with the starred symbol[1] $|\tilde{P}^\star\rangle$. It is easy to show that (see appendix B) $|\tilde{P}^\star\rangle$ is the right eigenvector with eigenvalue A_0 of

$$\tilde{H}^\star = -\alpha(a^+ - 1) - \delta(\mathrm{e}^s a^+ - 1) - \gamma(a_s^{\star -} - d_s^\star) - \beta(\mathrm{e}^{-s} a_s^{\star -} - d_s^\star), \tag{3.18}$$

where

$$a_s^{\star -} = \begin{pmatrix} 0 & w_{c,1,s} & 0 & 0 & \dots \\ 0 & 0 & w_{c,2,s} & 0 & \\ 0 & 0 & 0 & w_{c,3,s} & \\ 0 & 0 & 0 & 0 & \\ \vdots & & & & \ddots \end{pmatrix}, \tag{3.19}$$

$$a^+ = \begin{pmatrix} 0 & 0 & 0 & \dots \\ 1 & 0 & 0 & \\ 0 & 1 & 0 & \\ \vdots & & & \ddots \end{pmatrix}, \tag{3.20}$$

and the operator $d_s^\star$ has entries $\delta_{ij} w_{c,i,s}$. The operator $\tilde{H}^\star$ is equivalent to the s-modified Hamiltonian of a standard ZRP with departure rates $w_{c,n,s}$. However, it is not a genuine s-modified Hamiltonian for the on-off ZRP as it shares only the lowest eigenvalue A_0 with $\tilde{H}$ (the higher eigenvalues being different, in general) hence it only contains information about the limiting behaviour and does not generate the dynamics.

As a partial conclusion, we underline that both the systems with bounded and unbounded rates display the fluctuating behaviour seen in the standard ZRP as long as the ground state satisfies equations (3.15) and (3.16). This is certainly true for current fluctuations close to the mean $\langle j \rangle$. However, the effective interaction $w_{c,n,s}$ has a dependence on n and s different from the standard ZRP and this alters the range of validity of this regime. In the following, we show that larger current fluctuations in the on-off ZRP can be strongly affected by time correlations.

[1] Being defined in the occupation state space, $|\tilde{P}^\star\rangle$ is different from the vector $|\tilde{P}\rangle$ that is element of the joint occupation–clock space.

3.1.2 Dynamical phase transition

The scenario seen so far is an analytical continuation of the stationary state. Despite this, certain values of the bias s correspond to non-analyticity in the SCGF, as demonstrated in this subsection. From section 1.6 we recall that such a behaviour is often referred to as a dynamical phase transition because of the analogy of the SCGF with the Helmholtz Free Energy. A dynamical phase transition occurs as soon as one of the scalar products $\langle 1|\tilde{P}\rangle$ or $\langle \tilde{P}|P(0)\rangle$ diverges. It is worth noting that the choice of the initial distribution $|P(0)\rangle$ influences the value of the second norm. In order to ensure a finite $\langle \tilde{P}|P(0)\rangle$, we will always consider an empty site as initial condition, unless explicitly stated. We must also ensure that the norm $\langle \tilde{P}|\tilde{P}\rangle$ is finite, i.e., that the eigenvector is normalizable and the discrete eigenvalue A_0 exists. We now derive analytically the conditions under which the norms $\langle 1|\tilde{P}\rangle$ and $\langle \tilde{P}|\tilde{P}\rangle$ converge and it is possible to identify the SCGF with the lowest eigenvalue A_0 given by equation (3.12).

3.1.2.1 Linear departure rates

We focus first on the case with $\mu_n = n$. For this particular choice of the interaction, particles in the memory-less ZRP can never pile up and the current shows a smooth SCGF. On the contrary, in the on-off model, the particle blockade alters the statistics of small currents in the following way. During an off phase, particles accumulate on the site thus provoking a short-lived congestion, as the boundaries continue to inject particles. The presence of this mechanism suggests that the corresponding (short-lived) state cannot have the same structure as the NESS[2], hence the ansatz (3.3)–(3.5) is no longer adequate. From a mathematical point of view, a transition occurs when $\langle 1|\tilde{P}\rangle$ diverges. The condition $\langle 1|\tilde{P}\rangle < \infty$ is satisfied for $\lim_{n\to\infty} \rho_{n,s} < 1$ where $\rho_{n,s}$ is defined in equation (3.15). Taking its limit as $n \to \infty$, we get

$$\frac{(\beta+\gamma)\,(\alpha+\delta \mathrm{e}^{s})}{c(\beta+\gamma)+(\alpha+\delta \mathrm{e}^{s})\,(\gamma+\beta \mathrm{e}^{-s})} < 1. \tag{3.21}$$

For later convenience, we simplify this condition as

$$A_0 < c + \delta(1-\mathrm{e}^{s}), \tag{3.22}$$

[2] When the state turns on, the particles are released with a rate proportional to n—which can be very large—thus freeing the site from the congested particles and granting a steady state: the congestion is only instantaneous.

which is satisfied for $s > s_1$, where

$$\mathrm{e}^{s_1} = \frac{2\alpha\beta}{\alpha\beta - \beta\delta - \beta c - c\gamma + \sqrt{4\alpha\beta^2\delta + (\alpha\beta - \beta\delta - \beta c - c\gamma)^2}}. \tag{3.23}$$

In the PA case, s_1 is real and finite. However, when $\delta = 0$, the denominator in the argument of the logarithm is negative for c larger than the threshold value

$$c_0 = \frac{\alpha\beta}{\beta + \gamma}. \tag{3.24}$$

In other words, for $\delta = 0$ and $c > c_0$, the critical value s_1 is not well defined and the rare currents are not distinguishable from those of a standard ZRP with effective departure rates. This is particularly clear in the TA case, i.e. $(\gamma, \delta) = (0, 0)$, as the critical value is simply

$$\boxed{s_1 = \ln \frac{\alpha}{\alpha - c}} \tag{3.25}$$

which is well defined only for $c < c_0 = \alpha$.

We now prove that, when $\mu_n = n$, the norm $\langle \tilde{P} | \tilde{P} \rangle$ is always finite. In appendix C, the eigenvector $\langle \tilde{P}|$ is derived. Its components have a form similar to equations (3.3)–(3.5), with the factors $p_{\mathrm{ON},s}$ and ρ_s replaced by $p^{\mathrm{left}}_{\mathrm{ON},s} \equiv 1/2$ and $\rho^{\mathrm{left}}_s \equiv (\beta \mathrm{e}^{-s} + \gamma)/(\beta + \gamma)$ respectively. The series $\langle \tilde{P} | \tilde{P} \rangle$ is simplified by summing first the pairs corresponding to the same occupation number and the condition for convergence can be written as $\lim_{n\to\infty} \rho^{\mathrm{left}}_s \rho_{n,s} < 1$, which is always satisfied. Consequently, for linear departure rates, the only mechanism responsible for dynamical phase transitions is the on-off clockwork, which becomes dominant when $\langle 1 | \tilde{P} \rangle$ diverges.

It is also interesting to notice that, collapsing the components corresponding to $\tau = \mathrm{ON}$ and $\tau = \mathrm{OFF}$ for each n, we find the left eigenvector of the standard-ZRP Hamiltonian, i.e., $\langle \tilde{P}^\star | = \left(1, \rho^{\mathrm{left}}_s, \rho^{\mathrm{left}\,2}_s, \dots\right)$, where $\rho^{\mathrm{left}}_s = (\beta \mathrm{e}^{-s} + \gamma)/(\beta + \gamma)$. This is also the left eigenvector of the reduced operator (3.18).

3.1.2.2 Constant departure rates

Let us consider now the case $\mu_n = 1$, $n > 0$. The scalar product $\langle 1 | \tilde{P} \rangle$ is finite when the n-independent parameter ρ_s is less than 1 and a dynamical phase transition occurs at $\rho_s = 1$. In the PA case, the solution of this equation for s involves a cumbersome cubic and therefore is not reported here. Conversely, in the TA case, after some algebra we can

prove that the equality is verified at the critical value

$$\boxed{s_1 = \ln \frac{\alpha(\alpha - \mu\beta)}{(c\mu\beta - \alpha\mu\beta - c\alpha)}.} \tag{3.26}$$

In order to check whether the second norm $\langle \tilde{P}|\tilde{P}\rangle$ is finite, we again need the left eigenvector $\langle \tilde{P}|$. As the dependence on μ_n cancels in the left eigenproblem, $\langle \tilde{P}|$ is the same as the linear departure rate case (see again the appendix C). The condition for convergence is $\rho_s^{\text{left}}\rho_s < 1$ and the value of s such that $\rho_s^{\text{left}}\rho_s = 1$ is referred to as s_2. For the TA case, it has the closed form solution

$$\boxed{s_2 = -\ln \frac{(-c + \sqrt{c^2 + 4c\beta})}{2\alpha},} \tag{3.27}$$

while for the PA case, a closed form solution for $\rho_s^{\text{left}}\rho_s = 1$ has not been found. We underline that the values $s = s_1$ and $s = s_2$ mark the onsets of new phases.

It is important to notice that the scenario seen so far is entirely encoded into the operator $\tilde{H}^\star$ of equation (3.18). In fact, this operator not only has lowest eigenvalue A_0, as seen in section 3.1.1, but the normalisation of its ground-state eigenvector yields sums $\langle 1|\tilde{P}^\star\rangle$ and $\langle \tilde{P}^\star|\tilde{P}^\star\rangle$ that diverge at the same critical points as $\langle 1|\tilde{P}\rangle$ and $\langle \tilde{P}|\tilde{P}\rangle$, respectively.

3.1.3 Large current fluctuations

In this subsection, we focus on the large fluctuation regimes $s > s_1$ and $s < s_2$. We employ different approaches to study the large fluctuation regimes in the linear and constant departure rate cases.

3.1.3.1 Linear departure rates

To get some insights into the regimes with large current fluctuations we consider first a finite-dimensional version of the on-off ZRP on the single-site lattice. In fact, the SCGF on a discrete finite configuration space is always given by the smallest eigenvalue A_0 of the s-modified Hamiltonian, as the prefactors in (1.134) are always finite. For the TA case we truncate the Hamiltonian (3.1) by imposing a reflective boundary in the state

with occupation number N. The resulting matrix in block form is

$$H_{\mathsf{N}} = \left(\begin{array}{cc|cc|cc|c|cc} c+\alpha & 0 & 0 & 0 & 0 & 0 & \dots & & \\ -c & \alpha & 0 & -\beta\mu_1 & 0 & 0 & & & \\ \hline -\alpha & -\alpha & c+\alpha & 0 & 0 & 0 & & & \\ 0 & 0 & -c & \alpha+\beta\mu_1 & 0 & -\beta\mu_2 & \dots & & \\ \hline \vdots & & & & \vdots & \ddots & & \vdots & \\ & & & & & & 0 & -\beta\mu_{\mathsf{N}} \\ \hline & & & & & & c & 0 \\ & & & & & & \dots & -c & \beta\mu_{\mathsf{N}} \end{array}\right), \tag{3.28}$$

which defines a Master equation where the n-th block row specifies the dynamics of the configuration with occupation number n and within each block the first and the second rows correspond to an off and an on phase, respectively.

In the present linear departure rate case $\mu_n = n$ and the matrix H_{N} generates the dynamics of a special queue of type M/Ph/N/N, i.e., with Markovian arrival, phase-type departures, N servers and finite capacity N (see the Kendall notation in section 2.1). According to the procedure of section 1.6, the finite-capacity s-modified Hamiltonian $\tilde{H}_{\mathsf{N}}$ is obtained by multiplying the upper-diagonal rates μ_n ($n = 1, 2, 3, \dots, \mathsf{N}$) of H_{N} by e^{-s}.

The numerical evaluation of the spectrum of $\tilde{H}_{\mathsf{N}}$, see figure 3.1, shows that the two lowest eigenvalues get closer around s_1 with increasing values of N. This gives a clue about the limiting behaviour for $\mathsf{N} \to \infty$, where the eigenvalues would coalesce at $s = s_1$ and two different dynamical phases emerge. The SCGF converges to a constant branch of value c for $s > s_1$. An heuristic reasoning allows us to understand the mechanism that leads to this dynamical phase. In the limit $s \to \infty$, the truncated s-modified Hamiltonian is lower-diagonal and its eigenvalues are given by the escape rates. As long as the condition $c < c_0 = \alpha$ holds, the smallest eigenvalue is c; this is the escape rate from the configuration with N particles and off state. We argue that, in the infinite capacity case, such a configuration corresponds to an instantaneous situation where the occupation number grows indefinitely ("instantaneous" congested state) and clock is off. We also argue that, in the asymptotic limit as the time approaches infinity this is the dominant behaviour not only for $s \to \infty$, but also for the other values $s > s_1$. We expect that the corresponding eigenvector does not satisfy the same ansatz as the one of equations (3.15)–(3.16); hence the dynamical phase transition is due to the crossover of two distinct eigenvectors. This has been observed in spatially-extended non-equilibrium models, such as the Glauber model with open boundaries of reference [13].

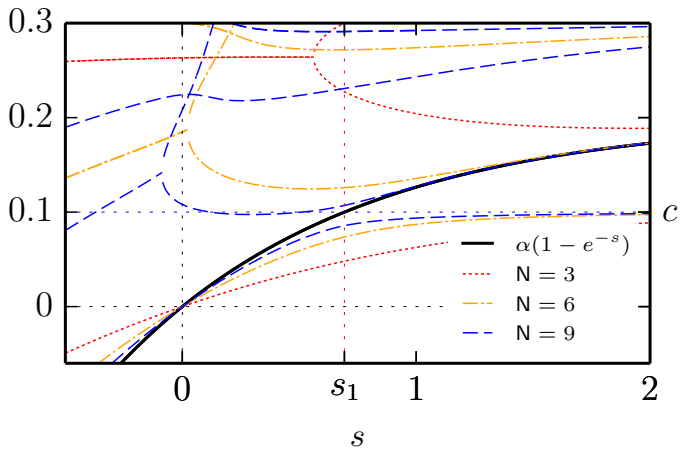

Figure 3.1: Real part of the spectrum of the finite-capacity s-modified Hamiltonian for parameters $(\alpha, \beta, \gamma, \delta, c) = (0.2, 0.3, 0, 0, 0.5)$ and $\mu_n = n$. The two lowest eigenvalues get closer at $s = s_1$ for increasing values of N and it seems that the merge in the infinite capacity limit $\mathsf{N} \to \infty$. For $s < s_1$ the smaller eigenvalue converges to $A_0 = \alpha(1 - \mathrm{e}^{-s})$, while for $s > s_1$ it converges to c.

Our prediction is checked against the output of the cloning algorithm of Lecomte and Tailleur [12] (see section 1.7.2 of this thesis). The results are shown in figure 3.2(a). Our implementation correctly reproduces the most relevant features of the SCGF, i.e., the non-analyticity at s_1 and the constant branch for $s > s_1$, but loses accuracy for large positive currents ($s < 0$) presumably due to the finite-ensemble effect.

In the PA process, the lowest eigenvalue does not appear to converge to a finite value in the limit $s \to \infty$. From the condition (3.22) for the eigenvalue crossover, we suggest

$$e(s) = \begin{cases} \frac{\alpha\beta}{\gamma+\beta}(1 - \mathrm{e}^{-s}) + \frac{\gamma\delta}{\gamma+\beta}(1 - \mathrm{e}^{s}), & s \leq s_1 \\ c + \delta(1 - \mathrm{e}^{s}). & s > s_1. \end{cases} \tag{3.29}$$

The right branch ($s > s_1$) can be physically understood by separating the contributions of the particles leaving the site rightwards, which contribute a term c as in the TA case, and the particles injected from the right boundary, which independently follow a Poisson process with rate δ and contribute a term $\delta(1-\mathrm{e}^{s})$. Since in this regime the particles pile up, the corresponding SCGF branch does not depend on the left boundary. Numerical simulations, shown in figure 3.2, confirm our argument. There is no analogue, in the memory-less ZRP, of the c-dependent dynamical phase for $s > s_1$, which arises as a consequence of the temporal correlations.

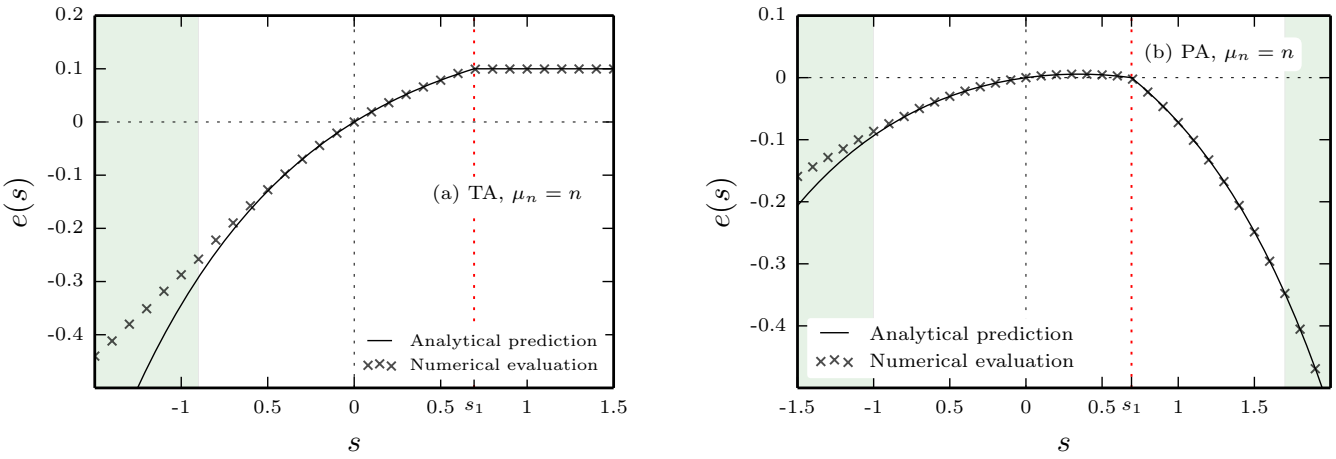

Figure 3.2: SCGF of the single-site on-off ZRP with $\mu_n = n$ for (a) TA hopping rates, $(\alpha, \beta, \gamma, \delta, c) = (0.2, 0.3, 0, 0, 0.1)$, and (b) PA hopping rates, $(\alpha, \beta, \gamma, \delta, c) = (0.1, 0.2, 0.1, 0.1, 0.1)$. Points are data from the cloning algorithm with number of clones $N = 10^4$ and simulation time $t = 10^4$. The SCGF is systematically overestimated for small values of s. This prescription is believed to be exact for $N \to \infty$, $t \to \infty$, and is not reliable when the cloning factor is larger than N (shaded areas). We expect a better approximation but a slow convergence for larger ensembles and longer simulation times.

For SCGFs with non differentiable points, as the one in equation (3.29), the Legendre–Fenchel transform (1.121) of $e(s)$ gives in general the convex hull of the true rate function which can hide a non-convex shape. However, for this system, we argue on physical grounds (see following) that equation (1.121) gives indeed the true rate function, which is

$$\hat{e}(j) = \begin{cases} c + \delta + j - j\ln(\frac{-j}{\delta}), & j \le j_{1,a} \\ -s_1 j + c + (1 - \mathrm{e}^{-s_1}), & j_{1,a} < j < j_{1,b} \\ \dfrac{\alpha\beta}{\beta+\gamma} + \dfrac{\gamma\delta}{\beta+\gamma} - \sqrt{j^2 + 4\dfrac{\alpha\beta}{\beta+\gamma}\dfrac{\gamma\delta}{\beta+\gamma}} + & \\ \quad + j\ln\dfrac{j + \sqrt{j^2 + 4\frac{\alpha\beta}{\beta+\gamma}\frac{\gamma\delta}{\beta+\gamma}}}{2\frac{\alpha\beta}{\beta+\gamma}}, & j \ge j_{1,b}. \end{cases} \tag{3.30}$$

The two critical currents

$$j_{1,a} = -\delta\,\mathrm{e}^{s_1}, \tag{3.31}$$

$$j_{1,b} = \frac{\alpha\beta}{\beta+\gamma}\,\mathrm{e}^{-s_1} - \frac{\gamma\delta}{\beta+\gamma}\,\mathrm{e}^{s_1} \tag{3.32}$$

are, respectively, the right and left derivatives of $e(s)$ at $s = s_1$. The value of $j_{1,a}$ is non-positive, and is zero in the TA process, the value of $j_{1,b}$ is strictly positive. The

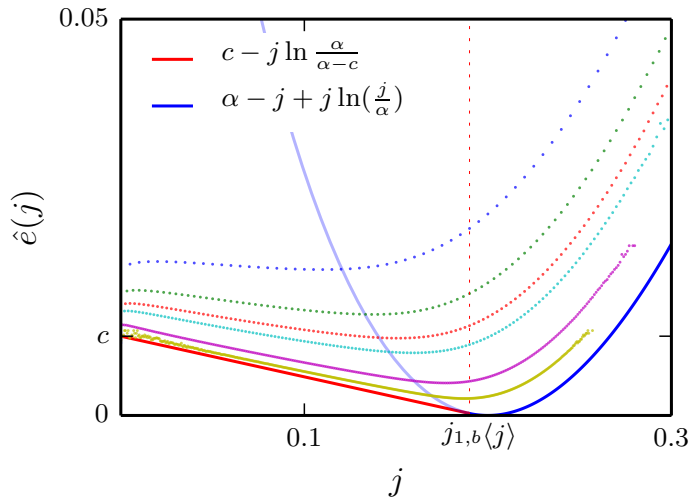

Figure 3.3: Rate function $\hat{e}(j)$ for the on-off TA process with $(\alpha, \beta, \gamma, \delta, c) = (0.2, 0.3, 0, 0, 0.01)$ and $\mu_n = n$ (solid blue line). The points are simulation data for the finite-time rate function $-\ln[\text{Prob}(J/t = j)]/t$ computed at $t = 200, 300, 400, 500, 600, 700$ (top to bottom). An ensemble size of $N = 10^{10}$ has been used. Notice that information about the tails of the rate function is quickly lost. The rate function of the memory-less ZRP for $s < s_1$ is plotted in light blue. Comparison with the solid red line manifests the extent of memory effects on rare events, while the statistics around $\langle j \rangle$ are identical.

phase $j \leq j_{1,a}$ is obtained from the Legendre–Fenchel transform of $e(s)$ in the interval $s > s_1$, while the phase $j \geq j_{1,b}$ is derived from $e(s)$, with $s < s_1$. The transition value s_1 is mapped to the whole linear branch in $j_{1,a} < j < j_{1,b}$. This behaviour is equivalent to an ordinary equilibrium first-order phase transition, where a linear branch of a thermodynamic potential corresponds to the coexistence of two phases. In this non-equilibrium system, the mixed phase consists in a regime where, for some finite fraction of time, the current assumes value $j_{1,a}$, while during the rest of the time it has value $j_{1,b}$. As a result, the rate function in this region is linear with j, as predicted by the Legendre–Fenchel transform. This argument is supported by standard Monte Carlo simulations and is particularly evident in the TA case, see figure 3.3. Such a figure, alongside figure 3.7, underlines the main difficulty in simulating rare trajectories; in fact, according to the large deviation principle, all the trajectories tend to display current close to the mean and obtaining information about the extreme events is hard, although the ensemble used is several orders of magnitude larger than the one used for the cloning simulations.

The different phases can be physically understood by observing the effect of the particle blockade. In the case with TA hopping rates, when the site is off, the particles

accumulate and the outgoing current is necessarily zero. The zero current is mapped to the flat section of the SCGF. This is the dominant mechanism responsible for zero current. At the end of an off period, we have a configuration with many particles on the site. When the lock is released, particles can leave the site with a rate proportional to the occupation number, the more the particles on site, the higher the departure rate. Consequently, the particles are quickly released after an off period and the current jumps to a positive value. In particular, the probability of having currents larger than $j_{1,b}$ is dominated by the phases during which the site is in a state on. In the presence of arrivals from the right boundary ($\delta \neq 0$), the blocked configuration becomes important for negative currents $j < j_{1,a}$, and the rate function has an additional term corresponding to an independent Poisson process with rate δ.

As an aside, the dynamical phase transition seen at s_1 is not restricted to the particular on-off ZRP explored here. For example, an alternative on-off ZRP with unbounded departure rates and on-off dynamics independent from the arrivals, displays the same fluctuating scenario. Specifically, we assume that the phase switches between the on and off states with rates respectively c and d and the total departure rate μ_n is proportional to n. The modified Hamiltonian is:

$$\tilde{H}' = \left(\begin{array}{cc|cc|cc|c} c+\alpha & -d & & & & & \cdots \\ -c & d+\alpha & & -\beta\mu_1 \mathrm{e}^{-s} & & & \\ \hline -\alpha & & c+\alpha & -d & & & \\ & -\alpha & -c & d+\alpha+\beta\mu_1 & & -\beta\mu_2 \mathrm{e}^{-s} & \\ \hline & & -\alpha & & c+\alpha & -d & \\ & & & -\alpha & -c & \alpha+\beta\mu_2+d & \\ \hline \vdots & & & & & & \ddots \end{array}\right). \tag{3.33}$$

The spectrum of the truncated version of $\tilde{H}'$ (with capacity N) reproduces the scenario of figure 3.4, which is similar to the one of figure 3.1. In the limit as $s \to \infty$, such a truncated Hamiltonian is block lower diagonal with the smallest eigenvalue $A_0 = \frac{1}{2}(c + d + \mu_{\mathsf{N}} - \mu_{\mathsf{N}}\sqrt{\frac{2(d-c)}{\mu_{\mathsf{N}}} + 1})$, which converges to c for $\mathsf{N} \to \infty$. Also, spatially-extended spin systems such as the contact process [12] and some kinetically constrained models [3, 4], can possess active and inactive phases coexisting at $s = 0$. For a review on kinetically constrained models see reference [15].

3.1.3.2 Constant departure rates

The study of the extreme fluctuation regimes when $\mu_n = 1$, $n > 0$, requires a different approach. In fact, in this case the spectrum of the operator $\tilde{H}$ has a continuous band

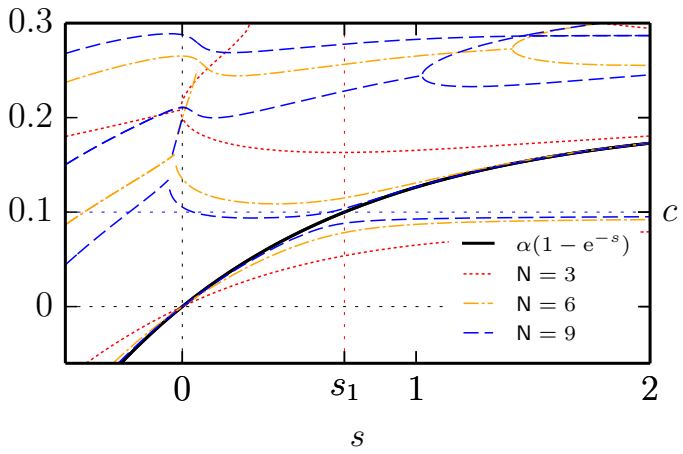

Figure 3.4: Real part of the spectrum of the finite-capacity version of s-modified Hamiltonian 3.33 for parameters $(\alpha, \beta, \gamma, \delta, c, d) = (0.2, 0.3, 0, 0, 0.5)$ and $\mu_n = n$. The minumum of its spectrum for $\mathsf{N} \to \infty$ is identical to one plotted in figure 3.1.

that governs the fluctuations in certain regimes (this can be inferred again by looking at the truncated operator). A way to obtain the SCGF is to evaluate the long-time limit of the matrix element $\langle 1|e^{-\tilde{H}t}|P(0)\rangle$ by computing the full spectrum and the complete set of eigenvectors of $\tilde{H}$ in the style of S. Karlin and J. L. McGregor [10] (see also a related discussion in section 1.2). This task appears to be rather complicated for the s-modified Hamiltonian (3.1), requiring spectral theory and integral representation of block non-stochastic operators[3]. As an approximation, we can use the reduced operator (3.18) and study the simpler expectation $\langle 1|e^{-\tilde{H}^\star t}|P(0)\rangle$. Recall that $\tilde{H}^\star$ has the same lowest eigenvalue A_0 as $\tilde{H}$, at least in the regime $s_2 \leq s \leq s_1$ where the ansatz (3.15)–(3.16) is valid. Outside this regime it is expected to yield only approximate information about the current fluctuations.

The integral representation allows us to take into account the dependence of the fluctuations on the initial condition. We follow the same procedure as in references [6, 14], with the difference that the departure rate here is $w_{n,c,s}$ and depends on s. Actually, it seems that the function $w_{n,c,s}$ only has a weak dependence on s for large positive or negative values of s (this can be checked numerically for many parameter combinations) but, nevertheless, we report the explicit calculations for completeness. As initial condition,

[3]The integral representation of Markov chains described by a *stochastic* block tridiagonal generator is derived, e.g., in reference [2] after some formidable mathematics.

we choose a geometric distribution with parameter x, i.e.,

$$|P(0)\rangle = (1-x)\sum_{n=0}^{\infty} x^n |e_n\rangle \tag{3.34}$$

where $|e_n\rangle$ denotes the configuration of the site with n particles and is an element of the natural basis for $\tilde{H}^\star$. The steady state is obtained for $x = z w_c^{-1}$, where

$$z = \frac{\alpha+\delta}{\beta+\gamma} \qquad \text{and} \tag{3.35}$$

$$w_c = \frac{\mu(\alpha+\delta+c)}{\alpha+\delta+c+(\beta+\gamma)\mu} \tag{3.36}$$

are the PA counterparts of the fugacity and effective departure rate, respectively, found in section 2.3, while the limit $x \to 0$ corresponds to the empty-site state.

The exact calculation of the full spectrum and of its eigenvectors, reported in appendix A, gives the following representation:

$$\begin{aligned}\langle 1|\mathrm{e}^{-\tilde{H}^\star t}|P(0)\rangle = &-\frac{1-x}{2\pi i x\phi}\oint_{C_1}\frac{\mathrm{e}^{-\varepsilon(\zeta)t}}{(\zeta-\frac{1}{x\phi})(\zeta-\frac{1}{\phi})}\,\mathrm{d}\zeta \\ &-\frac{1-x}{2\pi i x}\oint_{C_2}\frac{(y\zeta-1)\mathrm{e}^{-\varepsilon(\zeta)t}}{(\zeta-\frac{1}{\zeta\phi})(\zeta-\phi)(\zeta-y)}\,\mathrm{d}\zeta,\end{aligned} \tag{3.37}$$

where $\varepsilon(\zeta)$ is obtained from the expression for the continuous band of the spectrum $\epsilon(k)$ after the substitution $\zeta = \mathrm{e}^{ik}$ and

$$\phi = \sqrt{\frac{(\beta\mathrm{e}^{-s}+\gamma)w_{c,s}}{(\alpha+\delta\mathrm{e}^{s})}}, \tag{3.38}$$

$$y = \frac{1}{(\beta+\gamma)w_{c,s}}\sqrt{(\alpha+\delta\mathrm{e}^{s})(\beta\mathrm{e}^{-s}+\gamma)w_{c,s}}, \tag{3.39}$$

$$\epsilon(k) = \alpha+\delta+(\beta+\gamma)w_{c,s} - 2\sqrt{(\alpha+\delta\mathrm{e}^{s})(\beta\mathrm{e}^{-s}+\gamma)w_{c,s}}\cos(k). \tag{3.40}$$

The integration contours C_1 and C_2 are anti-clockwise circles centred around the origin with radius $\phi^{-1} < |\zeta| < (\phi x)^{-1}$ and infinitesimal size respectively.

The integrand of equation (3.37) has a saddle point at $\zeta = 1$. This suggests that we can compute the long-time limit of $\langle 1|\mathrm{e}^{-\tilde{H}^\star t}|P(0)\rangle$ by means of the method of steepest descents. In other words, we deform the integration contours C_1 and C_2 such that they pass through the saddle point. As the integrand also has poles on the real axis, care is needed when the new contour engulfs one of such poles and their residue must also be

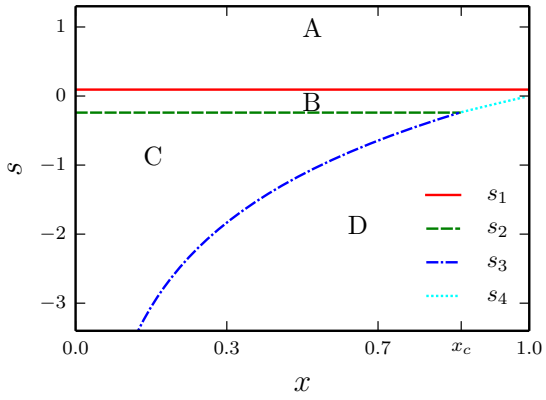

Figure 3.5: Phase diagram, based on equation (3.37), for the current fluctuations in the PA process with $\mu_n = 1$ and $(\alpha, \beta, \gamma, \delta, c) = (0.1, 0.2, 0.1, 0.1.0.5)$. The lines s_1 and s_4 correspond to first-order dynamical phase transitions while s_2 and s_3 mark second-order transitions. The red line $(s = s_1)$ corresponds to the solution of $\phi = y$. The green line $(s = s_2)$ corresponds to $y = 1$. The blue line $(s = s_3)$ and the cyan line $s = s_4$ are found from the solution of $1 = (\phi x)^{-1}$ and $y = (\phi x)^{-1}$ respectively. x_c is the point where $s_3 = s_4$.

taken into account[4].

For fixed parameters s and x, the leading term in the long time limit of $\langle 1|\mathrm{e}^{-\tilde{H}^* t}|P_0\rangle$ is given by the slowest decaying exponential and the SCGF is determined by one of the rates $\varepsilon(\phi), \varepsilon(y), \varepsilon(1), \varepsilon[(x\phi)^{-1}]$. Tuning s or x, the positions of the poles with respect to the saddle point contour are altered and, consequently, the leading term in the integral expansion changes. This produces the phase diagram of figure 3.5 for the SCGF. The critical line corresponding to the solution of $\phi = y$ is $s = s_1$. The line $s = s_2$ corresponds to $y = 1$. These two phase transitions were also previously found in section 3.1.2 as critical points in the joint occupation number and clock space. The curves $s = s_3$ and $s = s_4$ are solutions of $1 = (\phi x)^{-1}$ and $y = (\phi x)^{-1}$ respectively. The tri-critical point $s_3 = s_4$ is at x_c. It is worth noting that higher positive current fluctuations retain a dependence on the initial condition x and that, unlike the memory-less ZRP, the critical point s_1 can fall in the positive current range. While in the general PA case the curves that separate the dynamical phases do not seem to have closed form expressions, in the TA case, closed form solutions have been obtained with the help of symbolic computation software and reported in the following list (the colours refer to figure (3.5)):

(i) $s = s_1$ (red line). The knowledge of $|\tilde{P}^\star\rangle$ is sufficient to verify when the pre-factor

[4]The importance of the poles of partition function has been discussed, e.g., in reference [16] in the context of large deviation theory.

$\langle 1|\tilde{P}^\star\rangle$ is finite, i.e.,

$$\frac{\alpha(c+\mathrm{e}^{-s}\alpha+\beta\mu)}{(c+\mathrm{e}^{-s}\alpha)\beta\mu}<1, \tag{3.41}$$

$$\mathrm{e}^{-s}>\frac{-c\alpha+c\mu\beta-\alpha\mu\beta}{\alpha^2-\alpha\mu\beta}, \tag{3.42}$$

$$s<s_1=\ln\left(\frac{\alpha(\beta\mu-\alpha)}{c\alpha-c\beta\mu+\alpha\beta\mu}\right). \tag{3.43}$$

Notice that this condition makes sense when the denominator in the argument of the logarithm in (3.43) is positive, i.e., $c<\alpha\beta\mu/(\beta\mu-\alpha)$, while the stationarity condition $\alpha<\beta w_c$ ensures that the numerator is positive. This phase boundary can also be obtained from solving $\phi=y$.

(ii) $s=s_2$ (green line). This critical point marks the left boundary of the region where the condition $\rho_s^{\text{left}}\rho_s<1$ holds, i.e.,

$$\frac{\alpha(c+\mathrm{e}^{-s}\alpha+\beta\mu)}{(c+\mathrm{e}^{-s}\alpha)\beta\mu}\mathrm{e}^{-s}<1, \tag{3.44}$$

$$\mathrm{e}^{-s}<\frac{\sqrt{c^2+4c\mu}-c}{2\alpha}, \tag{3.45}$$

$$s>s_2=-\ln\left(\frac{\sqrt{c^2+4c\mu}-c}{2\alpha}\right). \tag{3.46}$$

It corresponds to a solution of $y=1$.

(ii) $s=s_3$ (blue line). This line corresponds to $(\phi x)^{-1}=1$. The critical point s_3 satisfies

$$\mathrm{e}^{-s_3}=\frac{\sqrt{(c\mu x^2+\alpha^2)^2+4\alpha^2\mu^2x^2}-c\mu x^2+\alpha^2}{2\alpha\mu x^2}. \tag{3.47}$$

(iv) $s=s_4$ (cyan line). This phase boundary is c-independent, specifically

$$\mathrm{e}^{-s_4}=x^{-1}. \tag{3.48}$$

It corresponds to the condition $y=(\phi x)^{-1}$.

The resulting phase diagram is similar to the PA one shown in figure 3.5, but with the transition lines identified by s_1 always mapped to a positive value of the current.

The four phases (in the general PA case) are detailed in the following list:

Phase A: $s>s_1$. In this case the leading term arises from the pole at $\zeta=\phi$. The

product $\langle 1|\tilde{P}\rangle$ diverges and the SCGF is different from the lowest eigenvalue A_0, being instead given by

$$e(s) = \delta(1 - \mathrm{e}^{s}) + \beta w_{c,s}(1 - \mathrm{e}^{-s}). \tag{3.49}$$

This phase corresponds to very small positive currents (if $\delta = 0$) or large backward currents (if $\delta > 0$). Large negative currents are mainly governed by the rate δ of particle arrival from the right, which contributes to the SCGF with the first term of (3.49). The second term corresponds to particles that jump rightwards from the site with an effective rate $\beta w_{c,s}$. The current fluctuations in this phase are optimally realised by a site with arbitrarily large occupation number that acts as a reservoir, so that the outgoing current has no dependence on the left boundary hops. This behaviour is equivalent to the instantaneous condensation seen in the linear departure rate case (subsection 3.1.3.1).

We argue that the presence of a left and a right term in equation (3.49) is generic for this phase, although there is no *a priori* reason for the effective rate $w_{c,s}$ to have the same form as in the small fluctuation regime. In the PA case, for large values of s, the SCGF is dominated by the first term and is not sensitive to the functional form of $w_{c,s}$.

Phase B: $(s_2 < s < s_1) \wedge (x < x_c)) \vee ((s_4 < s < s_1) \wedge (x > x_c))$. This phase arises when the pole at $\zeta = y$, corresponding to the lowest eigenvalue A_0 (3.12), becomes dominant, hence

$$e(s) = \frac{\alpha\beta}{\beta + \gamma}(1 - \mathrm{e}^{-s}) + \frac{\gamma\delta}{\beta + \gamma}(1 - \mathrm{e}^{s}). \tag{3.50}$$

The probability of fluctuations in this regime is asymptotically identical to the standard ZRP. In this range the site has finite occupation and the probability that a particle leaves is conditioned to an arrival event, just as in [5, 9, 14].

Phase C: $(x < x_c) \wedge (s_3 < s < s_2)$. This phase arises from the saddle-point at $\zeta = 1$. It corresponds to a large forward current sustained by a large inward current from the left boundary. Here the spectrum of $\tilde{H}^{\star}$ is completely continuous. The asymptotic form (1.134) still holds, as the prefactors $\langle 1|\tilde{P}_{A_0}\rangle$ and $\langle \tilde{P}_{A_0}|P(0)\rangle$ are finite, but with an oscillating (non-decaying in n) ground state $|\tilde{P}_{A_0}\rangle$. This also represents an instantaneous condensate, but with particle number growing as the square root of time [14]. Here the spectrum of $\tilde{H}^{\star}$ is continuous and the SCGF is given by its minimum (3.40):

$$e(s) = \epsilon(0) = \alpha + \delta + (\beta + \gamma)w_{c,s} - 2\sqrt{(\alpha + \delta\mathrm{e}^{s})(\beta\mathrm{e}^{-s} + \gamma)w_{c,s}}. \tag{3.51}$$

Phase D: $[(s < s_3) \wedge (x < x_c)] \vee [(s < s_4) \wedge (x > x_c)]$. This phase arises when the

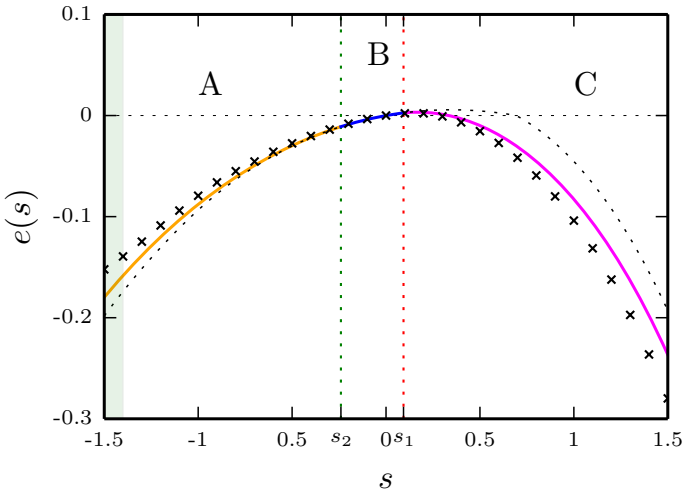

Figure 3.6: SCGF of the on-off ZRP with $(\alpha, \beta, \gamma, \delta, c) = (0.1, 0.2, 0.1, 0.1, 0.5)$ and $\mu_n = 1$. Points are data from the cloning simulations, $N = 10^4, t = 10^4$. Dotted line is the SCGF of the Markovian-ZRP ($c \to \infty$) with same boundary rates. Solid line is the analytic approximation (3.49)–(3.51). The SCGF of the ZRP with s-independent departure rate w_c would overlap the solid line at this scale.

residue at $\zeta = (\phi x)^{-1}$ dominates the long-time behaviour:

$$e(s) = \alpha + \delta + (\beta + \gamma)w_{c,s} - (\beta \mathrm{e}^{-s} + \gamma)w_{c,s}x - (\alpha + \delta \mathrm{e}^{s})/x. \tag{3.52}$$

It corresponds to a large forward current of particles that is most likely to be realized from an initial configuration with very high occupation number and also has an analogue in the standard ZRP [14].

These results are compared to the cloning simulations in figure 3.6 for $x \to 0$. Similarly to the independent-particle case, the cloning data for the left branch, corresponding to large positive currents, is potentially affected by finite ensemble effects, as documented in reference [8]. It turns out that for the chosen parameters our approximation (3.49)–(3.51), plotted as a solid line, is very close to the naive approach (not shown) in which the same representation (3.37) is used, but the effective departure rate has the s-independent form w_c (defined in section 2.3) for all the regimes. The analytical SCGF does not match the simulation points in both phases A and C. We attribute this to the failure of the assumption (3.15)–(3.16) for the ground state in phases A and C. In other words, large fluctuations cannot be exactly described by an effective departure rate $w_{c,s}$ with a simple functional dependence on s.

As a last step, in figure 3.7, the rate function $\hat{e}(j)$, computed by means of a Legendre–

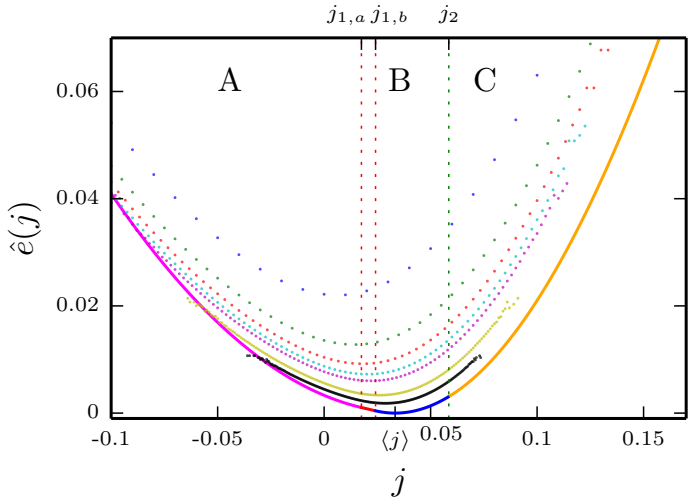

Figure 3.7: Rate function for the on-off ZRP. Parameter combinations as in figure 3.6. Points are data for $-\ln[\text{Prob}(J/t = j)]/t$ from standard Monte Carlo simulations at times $t = 100, 200, 300, 400, 500, 1000, 2000$ (top to bottom) with an ensemble size $N = 10^{10}$. The solid line is the analytical result for $t \to \infty$; in the central range it is the exact asymptotic limit, while in the regimes A and C the line is an approximated solutions, as discussed in subsection 3.1.1.1, 3.1.2.2, and 3.1.3.2.

Fenchel transform on the SCGF (3.49)–(3.51), is compared to the finite-time rate functions obtained from standard Monte Carlo simulations with an ensemble size of 10^{10}. Although approximate, $\hat{e}(j)$ appears to capture well the shape of the long time limit for the simulation data points.

3.2 Numerical results for the many-site system

In chapter 2 we showed that the on-off ZRP on an extended lattice lacks a product-form stationary solution. This, in turn, makes the analytical study of fluctuations, across the generic bond, impractical. It would be possible to use the mean-field stationary solution to derive an approximate SCGF using the same procedure as in the single-site model. However, we do not expect the result to be accurate, especially for small values of c and for current fluctuations far from the mean. To explore the larger system we make use of the cloning method. This small section is concerned with the full statistics of the empirical currents j_l, which is the difference between the number of particle hops from site l to site $l+1$ and the number of hops from site $l+1$ to site l, divided by the observation time t. This definition is extended to the input current j_0 and output current j_L. In

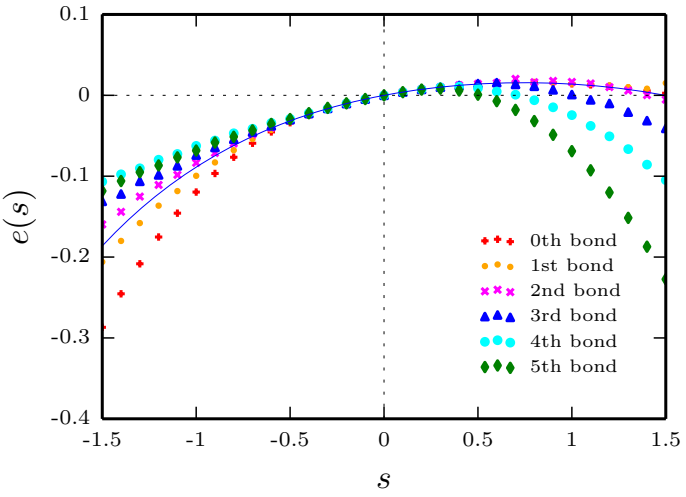

Figure 3.8: Simulation results for the SCGF in a five-site on-off ZRP with $\mu_n = 1$ and parameters $(\alpha, \beta, \gamma, \delta, p, q, c) = (0.1, 0.2, 0.1, 0.1, 0.55, 0.45, 0.5)$. The solid line is the c-independent lowest eigenvalue of the s-modified Hamiltonian for the five-site Markovian-ZRP.

figure 3.8 the results of the cloning simulation for a one-dimensional lattice are plotted. While the statistics of rare currents is bond-dependent, it is possible to appreciate that, for each bond, the SCGF matches that of a Markovian ZRP in the neighbourhood of $s = 0$, a feature shared with the one-site system. The SCGF of the Markovian ZRP for the one-dimensional lattice has been obtained in Rákos and Harris [14] and is reported here for convenience

$$e(s) = \alpha'(1 - \mathrm{e}^{-s}) + \delta'(1 - \mathrm{e}^{s}), \tag{3.53}$$

where

$$\alpha' = \frac{(p-q)(p/q)^L q\alpha\beta}{p(p-q-\beta)\gamma + (p/q)^L q\beta(p-q+\gamma)}, \tag{3.54}$$

$$\delta' = \frac{p(p-q)\gamma\delta}{p(p-q-\beta)\gamma + (p/q)^L q\beta(p-q+\gamma)}. \tag{3.55}$$

As a general consequence, the central regime satisfies a Gallavotti–Cohen fluctuation symmetry [11] with $E = \ln[(p/q)^{L-1}\alpha\beta/(\gamma\delta)]$ (see also section 1.4.2). Such a symmetry has been discussed in references [5, 6, 14] for the functional form (3.53). In the on-off ZRP, it seems to be ensured by the fact that the relative probabilities of particle jumps towards the left or the right are independent of the time that the particle spends on a site; this property is indeed the direction-time independence seen in section 4.3, which has been proved (in reference [1]) to be a sufficient condition for the fluctuation symmetry to

hold in finite state space. However, the fluctuation symmetry is not guaranteed to hold on an arbitrary domain in systems with infinite state space [6]. In fact, as expected, we see here a c-dependent breakdown for large fluctuations.

3.3 Discussion

In this chapter we have supplemented the stationary-state analysis for the on-off ZRP of the previous chapter with an investigation of the particle current fluctuations in the same model. The selected tool to deal with such fluctuations is the large deviation theory. We mainly studied, both numerically and analytically, the one-node system. Fluctuations around the mean are obtained by analytic continuation of the stationary state and are indistinguishable from those of a memory-less ZRP. However, under certain conditions, large current fluctuations are optimally realized by the instantaneous piling up of particles on the site and the statistics of such fluctuations change abruptly.

In the absence of direct inter-particle interaction we have found a memory-induced dynamical first-order phase transition, i.e., the SCGF $e(s)$ is non-analytic at a particular value s_1. In the totally asymmetric case, this occurs only if the parameter c is smaller than the arrival rate α. The system with constant departure rates, i.e., attractive inter-particle interaction, undergoes second-order as well as first-order dynamical phase transitions. The state of the system during a small fluctuation event has the same form as the stationary state, but with a more general modified effective interaction factor. Indeed, the exact phase boundaries and the large deviation function of this regime are encoded in the reduced operator $\tilde{H}^\star$ of equation (3.18), which has the same structure as the s-modified Hamiltonian of the standard ZRP, but with an s-dependent effective interaction factor. We have used the same operator $\tilde{H}^\star$ to find an approximate solution for the fluctuations outside this phase. Numerical tests confirm the presence of the predicted c-dependent dynamical phase transitions.

The separation between a small-fluctuation regime, with a memory independent SCGF, and high-fluctuation regimes, where memory plays a more obvious role, is a feature also found numerically in the spatially-extended system. It would be of interest to explore the role of topology in more detail as well as to look for similar memory effects in other driven interacting-particle systems and complex systems in general.

To conclude, these results highlight a situation where complex interactions are not seen in the stationary state and do not affect, qualitatively, the typical long-time behaviour of a system. They emerge at the fluctuating level and alter the probability of

observing rare phenomena. Such an observation leaves us with an interesting open question: "it is possible to predict when theories of real-world complex systems fail to forecast their rare behaviour, even though they are in perfect agreement with the typical one?". To attempt an answer to such a question, we need to be able to explore the large deviations of non-Markovian non-equilibrium systems *systematically*. As analytical progress is difficult (this chapter indeed revealed a number of subtleties) it would be important to resort to and rely on numerical methods appropriate for non-Markovian systems. This is the topic of the next chapter.

Bibliography

[1] D. Andrieux and P. Gaspard. The fluctuation theorem for currents in semi-Markov processes. *J. Stat. Mech.*, 2008(11):P11007, 2008.

[2] H. Dette, B. Reuther, W. J. Studden, and M. Zygmunt. Matrix measures and random walks with a block tridiagonal transition matrix. *SIAM J. Matrix Anal. Appl.*, 29(1):117–142, 2007.

[3] J. P. Garrahan, R. L. Jack, V. Lecomte, E. Pitard, K. van Duijvendijk, and F. van Wijland. Dynamical first-order phase transition in kinetically constrained models of glasses. *Phys. Rev. Lett.*, 98(19):195702, 2007.

[4] J. P. Garrahan, R. L. Jack, V. Lecomte, E. Pitard, K. van Duijvendijk, and F. van Wijland. First-order dynamical phase transition in models of glasses: an approach based on ensembles of histories. *J. Phys. A: Math. Gen.*, 42(7):075007, 2009.

[5] R. J. Harris, A. Rákos, and G. M. Schütz. Current fluctuations in the zero-range process with open boundaries. *J. Stat. Mech.*, 2005(08):P08003–P08003, 2005.

[6] R. J. Harris, A. Rákos, and G. M. Schütz. Breakdown of Gallavotti–Cohen symmetry for stochastic dynamics. *EPL*, 75(2):227–233, 2006.

[7] R. J. Harris, V. Popkov, and G. M. Schütz. Dynamics of instantaneous condensation in the ZRP conditioned on an atypical current. *Entropy*, 15(11):5065–5083, 2013.

[8] P. I. Hurtado and P. L. Garrido. Current fluctuations and statistics during a large deviation event in an exactly solvable transport model. *J. Stat. Mech.*, 2009(02): P02032, 2009.

[9] R. Juhász, L. Santen, and F. Iglói. Partially asymmetric exclusion models with quenched disorder. *Phys. Rev. Lett.*, 94(1):010601, 2005.

[10] S. Karlin and J. L. McGregor. The differential equations of birth-and-death processes, and the Stieltjes moment problem. *T. Am. Math. Soc.*, 85(2):489–489, 1957.

[11] J. L. Lebowitz and H. Spohn. A Gallavotti–Cohen type symmetry in the large deviation functional for stochastic dynamics. *J. Stat. Phys.*, 95(1/2):333–365, 1999.

[12] V. Lecomte and J. Tailleur. A numerical approach to large deviations in continuous time. *J. Stat. Mech.*, 2007(03):P03004–P03004, 2007.

[13] S. R. Masharian, P. Torkaman, and F. H. Jafarpour. Particle-current fluctuations in a variant of the asymmetric Glauber model. *Phys. Rev. E*, 89(1):012133, 2014.

[14] A. Rákos and R. J. Harris. On the range of validity of the fluctuation theorem for stochastic Markovian dynamics. *J. Stat. Mech.*, 2008(05):P05005, 2008.

[15] F. Ritort and P. Sollich. Glassy dynamics of kinetically constrained models. *Adv. Phys.*, 52(4):219–342, 2003.

[16] H. Touchette, R. J. Harris, and J. Tailleur. First-order phase transitions from poles in asymptotic representations of partition functions. *Phys. Rev. E*, 81(3):030101, 2010.

4 A numerical approach to large deviations in non-Markovian processes

Contents

As we have already seen, when the stochastic dynamics of a model system are Markovian, i.e., memory-less, we can specify the rules for its evolution in time by means of the constant rates g_{x_i,x_j} of transitions from configuration x_i to configuration x_j. The full set of rates encodes inter-event times with simple exponential waiting-time distributions (WTDs), which indeed possess the memory-less property. However, to model real-world systems, there is little double that such a naive description may not be appropriate. In fact, non-exponential WTDs and non-Markovian dynamics seem to be the norm in

many contexts, e.g., in physics [4], finance [44], biology [33], teletraffic engineering [40], and complex systems in general [20]. Furthermore, reckoning with non-Markovian dynamics is essential in non-equilibrium statistical mechanics. In fact, a strategy to study non-equilibrium systems is to target trajectories in space–time rather than static configurations [17, 32]; to this purpose, the details of the time evolution of the system, including the presence of non-exponential WTDs, are essential.

The model of chapters 2 and 3 incorporates memory by means of a device, called phase or clock, that stores information about past events (specifically particle arrivals). As seen in the previous chapters, this device permits analytical and numerical progress using tools developed for Markovian processes. However, there are processes whose waiting times cannot be reconstructed by means of the inclusion of hidden phases, as their WTDs simply cannot be expressed in terms of such phases (see the discussion on semi-Markov processes in section 1.3). In an even more general case, the waiting times depend on the whole history, as anticipated in section 1.5. Such genuine non-Markovian models are widespread in physics literature [7, 8, 37, 38, 43], but analytical progress is difficult and simulations are necessary to explore them systematically, especially at the fluctuating level. However numerical schemes able to efficiently probe large deviation functionals have been discussed only for memory-less systems [18, 22, 29, 34, 35].

Based on our results in Cavallaro and Harris [10], we present here a general numerical method to generate non-Markovian trajectories corresponding to arbitrarily rare currents, which is a generalisation of the cloning procedures of Giardinà et al. [18] and Lecomte and Tailleur [29]. This chapter is organised as follows. In section 4.1 we briefly recall the general formalism (making explicit the dependence on the history). In section 4.2 we present the simulation scheme for non-Markovian processes and the numerical method to evaluate their large deviation functionals. Section 4.3 deals with semi-Markov processes, where the formalism has a particularly lucid interpretation in terms of the *generalised Master equation*. In section 4.4 we test the method in some examples of increasing complexity, where the large deviation functions can be computed exactly and hence serve to test the exactness of our numerical method. We conclude the chapter in section 4.5.

4.1 Thermodynamics of trajectories

Following the argument of section 1.6, we are interested in the partition function

$$Z(s,t) = \int \mathrm{e}^{-sJ[w(t)]} \varrho[w(t)] \mathrm{d}w(t), \tag{1.119}$$

where the probability density of observing a sample path starting at t_0 and ending at time t is given, in the most general case, by equation (1.115). In this chapter we consider observables of type A, whose partition function can be written explicitly as

$$Z(s,t) = \sum_{n=0}^{\infty} \int_{t_0}^{t} \mathrm{d}t_1 \int_{t_1}^{t} \mathrm{d}t_2 \ldots \int_{t_{n-1}}^{t} \mathrm{d}t_n \sum_{x_0,x_1,\ldots,x_n} \phi_{x_n}[t-t_n; w(t_n)]\tilde{\psi}_{x_n,x_{n-1}}[t_n - t_{n-1}; w(t_{n-1})] \times \ldots \times \tilde{\psi}_{x_1,x_0}[t_1 - t_0; w(t_0)]P_{x_0}(t_0), \quad (4.1)$$

where

$$\tilde{\psi}_{x_n,x_{n-1}}[t_n - t_{n-1}; w(t_{n-1})] = \mathrm{e}^{-s\theta_{x_n,x_{n-1}}} \psi_{x_n,x_{n-1}}[t_n - t_{n-1}; w(t_{n-1})] \quad (4.2)$$

is the "biased" WTD of a stochastic dynamics that does not conserve total probability. The strategy is to simulate an ensemble of trajectories and measure the behaviour of in the long time limit. The exponential rate of divergence of equation (4.1) in the limit as $T = t - t_0$ approaches infinity gives the SCGF

$$e(s) = -\lim_{T\to\infty} \frac{1}{T} \ln Z(s,t), \quad (4.3)$$

which in turns can be used to compute the rate function

$$\hat{e}(j) = \sup_s \{e(s) - s\, j\}, \quad (4.4)$$

as we have already seen in section 1.6.

4.2 The "cloning" approach to large deviations in non-Markovian processes

4.2.1 Non-Markovian stochastic simulation

The recent interest in non-Markovian processes can be seen in a number of publications dealing with efficient methods to simulate their dynamics, either exact [1, 9, 12] or approximated [6, 13, 31]. Here we present a naive exact scheme to generate sample paths of non-Markovian processes, which provides the backbone for the non-Markovian cloning method. This outputs a series of waiting times and instantaneous configuration changes, consistent with the Markov and semi-Markov cases detailed in chapter 1. Generically, the WTD can be expressed in terms of a time-dependent *rate* or *hazard* $g_{x_n,x_{n-1}}[t_n - t_{n-1}; w(t_{n-1})]$, which is the probability density that there is a jump from x_{n-1} to x_n in

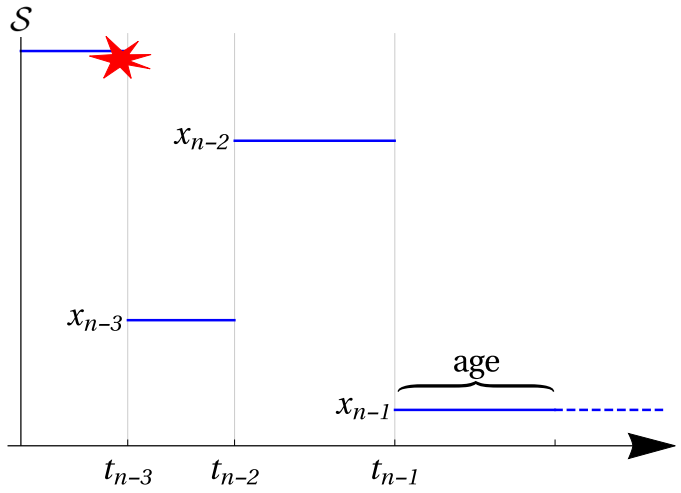

Figure 4.1: Representation of a portion of trajectory. The time elapsed from the last jump is called age. The conditional probability of having a configuration x_n at an instant $t_n > t_{n-1}$ depends on the age, as well as on events which happened during the history (e.g., the one marked by the red star).

$[t_n, t_n + \mathrm{d}t)$, conditioned on having no transitions during the interval $[t_{n-1}, t_n)$,

$$\psi_{x_n,x_{n-1}}[t_n - t_{n-1}; w(t_{n-1})] = g_{x_n,x_{n-1}}[t_n - t_{n-1}; w(t_{n-1})] \times \phi_{x_{n-1}}[t_n - t_{n-1}; w(t_{n-1})]. \tag{4.5}$$

For brevity we define $\tau = t_n - t_{n-1}$, which is the value of the *age*, i.e. the time elapsed since the last jump, when the next jump takes place. This is illustrated in figure 4.1. Roughly speaking, the hazard $g_{x_n,x_{n-1}}[\tau; w(t_{n-1})]$ is the likelihood of having an almost immediate transition from a state x_{n-1} known to be of age τ, to a state x_n, and, crucially, can also depend on the history $w(t_{n-1})$. From equation (4.5), summing over $x_n \in \mathcal{S}$, and defining $g_{x_{n-1}}[\tau; w(t_{n-1})] = \sum_{x_n} g_{x_n,x_{n-1}}[\tau; w(t_{n-1})]$ and $\psi_{x_{n-1}}[\tau; w(t_{n-1})] = \sum_{x_n} \psi_{x_n,x_{n-1}}[\tau; w(t_{n-1})]$, we get

$$\psi_{x_{n-1}}[\tau; w(t_{n-1})] = -\frac{\mathrm{d}}{\mathrm{d}\tau}\phi_{x_{n-1}}[\tau; w(t_{n-1})] \tag{4.6}$$

$$= g_{x_{n-1}}[\tau; w(t_{n-1})] \times \phi_{x_{n-1}}[\tau; w(t_{n-1})]. \tag{4.7}$$

The age-dependent sum $g_{x_{n-1}}[\tau; w(t_{n-1})]$ is also referred to as the *escape rate* from x_{n-1}, and using equation (4.7), can be written as a logarithmic derivative

$$g_{x_{n-1}}[\tau; w(t_{n-1})] = -\frac{\mathrm{d}}{\mathrm{d}\tau} \ln \phi_{x_{n-1}}[\tau; w(t_{n-1})]. \tag{4.8}$$

Integrating equation (4.8) with initial condition $\phi_{x_{n-1}}[0; w(t_{n-1})] = 1$ gives:

$$\phi_{x_{n-1}}[\tau; w(t_{n-1})] = \exp\left(-\int_0^\tau g_{x_{n-1}}[t; w(t_{n-1})]\,\mathrm{d}t\right), \tag{4.9}$$

$$\psi_{x_{n-1}}[\tau; w(t_{n-1})] = g_{x_{n-1}}[t; w(t_{n-1})]\exp\left(-\int_0^\tau g_{x_{n-1}}[t; w(t_{n-1})]\,\mathrm{d}t\right). \tag{4.10}$$

Equation (4.5) can be cast in the form (1.40), which is convenient for simulations,

$$\psi_{x_n,x_{n-1}}[\tau; w(t_{n-1})] = p_{x_n,x_{n-1}}[\tau; w(t_{n-1})] \times \psi_{x_{n-1}}[\tau; w(t_{n-1})], \tag{4.11}$$

where

$$p_{x_n,x_{n-1}}[\tau; w(t_{n-1})] = \frac{g_{x_n,x_{n-1}}[\tau; w(t_{n-1})]}{\sum_{x_n} g_{x_n,x_{n-1}}[\tau; w(t_{n-1})]}, \tag{4.12}$$

is the probability that the system jumps into the state x_n, given that it escapes the state x_{n-1} at age τ. It is important to notice that the normalization conditions

$$\sum_{x_n} p_{x_n,x_{n-1}}[\tau; w(t_{n-1})] = 1, \tag{4.13}$$

$$\int_0^\infty \psi_{x_{n-1}}[\tau; w(t_{n-1})]\,\mathrm{d}\tau = 1 \tag{4.14}$$

are satisfied. Hence, we can sample a random waiting time τ, according to the density $\psi_{x_{n-1}}[\tau; w(t_{n-1})]$ and, after that, a random arrival configuration x_n, according to the probability mass $p_{x_n,x_{n-1}}[\tau; w(t_{n-1})]$. This suggests the standard Monte Carlo algorithm for the generation of a trajectory $w(t)$ (1.115):

1) Initialise the system to a configuration x_0 and a time t_0. Set a counter to $n = 1$.
2) Draw a value τ according to the density (4.10) and update the time to $t_n = t_{n-1}+\tau$.
3) Update the system configuration to x_n, with probability given by (4.12).
4) Update n to $n + 1$ and repeat from 2) until t_n reaches the desired simulation time $T = t - t_0$.

4.2.2 The cloning step

We now need to take into account the effect of the factor $\mathrm{e}^{-s\theta_{x_n,x_{n-1}}}$ on the dynamics, which is to increment (if $\theta_{x_n,x_{n-1}} < 0$) or decrement (if $\theta_{x_n,x_{n-1}} > 0$) the "weight" of a trajectory, within an ensemble. This can be implemented by means of the cloning method, introduced in section 1.7.2. Here, we refine and extend this idea to the case of non-Markovian processes. One of the devices used in the original literature on Markovian cloning methods [18, 29] is to define the modified transition probabilities (1.141) and

(1.142), which are indeed valid *only* under the Markovian assumption, and a modified cloning factor (1.143), encoding the contraction or expansion of the trajectory weight. In fact, it is implicit in the original work that the redefinition of such quantities is unnecessary; in some cases it may also be inconvenient, see section 4.3.2. An arguably more natural choice, especially for non-Markovian dynamics, is to focus on the WTDs. Specifically, equations (4.1) and (4.2) suggest the following procedure:

1) Set up an ensemble of N clones and initialise each with a given time t_0, a random configuration x_0, and a counter $n = 0$. Set a variable C to zero. For each clone, draw a time τ until the next jump from the density $\psi_{x_0}[\tau; w(t_0)]$ and then choose the clone with the smallest value of $t = t_0 + \tau$.

2) For the chosen clone, update n to $n + 1$ and then the configuration from x_{n-1} to x_n according to the probability mass $p_{x_n,x_{n-1}}[\tau; w(t - \tau)]$.

3) Generate a new waiting time τ for the updated clone according to $\psi_{x_n}[\tau; w(t)]$ and increment its value of t to $t + \tau$.

4) **Cloning step.** Compute $y = \lfloor \mathrm{e}^{-s\theta_{x_n,x_{n-1}}} + u \rfloor$, where u is drawn from a uniform distribution on $[0, 1)$.

    1) If $y = 0$, prune the current clone. Then replace it with another one, uniformly chosen among the remaining $N - 1$.
    2) If $y > 0$, produce y copies of the current clone. Then, prune a number y of elements, uniformly chosen among the existing $N + y$.

5) Increment C to $C + \ln[(N + \mathrm{e}^{-s\theta_{x_n,x_{n-1}}} - 1)/N]$. Choose the clone with the smallest t, and repeat from 2) until $t - t_0$ for the chosen clone reaches the desired simulation time T.

The SCGF is finally recovered as $-C/T$ for large T. The net effect of step 4) is to maintain a constant population of samples whose mean current does not decay to $\langle j \rangle$.

4.3 Semi-Markov systems

We now focus on the statistics of time-extensive variables for semi-Markov processes based on references [2, 16, 30].

4.3.1 s-modified generalised Master equation

Recall that, in systems described by a standard Master equation, one strategy is to analyse a process that obeys a modified rate equation, obtained replacing the time-independent rates[1] g_{x_i,x_j}, with the products $\mathrm{e}^{-s\theta_{x_i,x_j}} g_{x_i,x_j}$, which are referred to as "biased" rates (as seen in chapter 1). In semi-Markov systems it is possible to investigate the statistics of $J[w(t)]$ in a similar, but more general, way. Instead of the standard Master equation, we deploy the GME (1.84). The probability $P_{(x_i,\mathsf{J})}(t)$ of having a configuration x_i with total current J at time t, under the constraint that the current can only grow or decrease by one unit at each jump, obeys the following GME:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}P_{(x_i,\mathsf{J})}(t) &= I_{(x_i,\mathsf{J})}(t-t_0) + \sum_{x_j\neq x_i}\int_{t_0}^{t} K_{(x_i,\mathsf{J})\leftarrow(x_j,\mathsf{J})}(t-\tau)P_{(x_j,\mathsf{J})}(\tau)\,\mathrm{d}\tau \\ &+\sum_{x_j}\int_{t_0}^{t} K_{(x_i,\mathsf{J})\leftarrow(x_j,\mathsf{J}+1)}(t-\tau)P_{(x_j,\mathsf{J}+1)}(\tau)\,\mathrm{d}\tau+\sum_{x_j}\int_{t_0}^{t} K_{(x_i,\mathsf{J})\leftarrow(x_j,\mathsf{J}-1)}(t-\tau)P_{(x_j,\mathsf{J}-1)}(\tau)\,\mathrm{d}\tau \\ &- \sum_{x_j\neq x_i}\int_{t_0}^{t} K_{(x_j,\mathsf{J})\leftarrow(x_i,\mathsf{J})}(t-\tau)P_{(x_i,\mathsf{J})}(\tau)\,\mathrm{d}\tau - \sum_{x_j}\int_{t_0}^{t} K_{(x_j,\mathsf{J}+1)\leftarrow(x_i,\mathsf{J})}(t-\tau)P_{(x_i,\mathsf{J})}(\tau)\,\mathrm{d}\tau \\ &- \sum_{x_j}\int_{t_0}^{t} K_{(x_j,\mathsf{J}-1)\leftarrow(x_i,\mathsf{J})}(t-\tau)P_{(x_i,\mathsf{J})}(\tau)\,\mathrm{d}\tau. \quad (4.15)\end{aligned}$$

The reader is reminded that $I_{(x_i,\mathsf{J})}(t-t_0)$ is an initial-condition dependent term, which contains the primed WTDs corresponding to the first jump of a sample path. We now make the assumption that the memory kernels are independent of the time-integrated current (only depending on the current increment), i.e.,

$$K_{(x_i,\mathsf{J})\leftarrow(x_j,\mathsf{J}-c)}(t) = K_{x_i,x_j,c}(t), \quad (4.16)$$

where $c=-1,0,1$. The system is diagonalised with respect to the current subspace by means of the discrete Laplace transform

$$\tilde{P}_{x_i}(t) = \sum_{\mathsf{J}} \mathrm{e}^{-s\mathsf{J}} P_{(x_i,\mathsf{J})}(t) \quad (4.17)$$

[1] In a slight shift in notation, we now use x_j and x_i as configuration labels.

and is then equivalent to

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\tilde{P}_{x_i}(t) = \tilde{I}_{x_i}(t-t_0) + \sum_{x_j\neq x_i}\int_{t_0}^{t} K_{x_i,x_j,0}(t-\tau)\tilde{P}_{x_j}(\tau)\,\mathrm{d}\tau \\ + \sum_{x_j}\int_{t_0}^{t} \mathrm{e}^{s} K_{x_i,x_j,-1}(t-\tau)\tilde{P}_{x_j}(\tau)\mathrm{d}\tau + \sum_{x_j}\int_{t_0}^{t} \mathrm{e}^{-s} K_{x_i,x_j,+1}(t-\tau)\tilde{P}_{x_j}(\tau)\,\mathrm{d}\tau \\ - \sum_{c,x_j\neq x_i}\int_{t_0}^{t} K_{x_j,x_i,c}(t-\tau)\tilde{P}_{x_i}(\tau)\,\mathrm{d}\tau - \int_{t_0}^{t} K_{x_i,x_i,-1}(t-\tau)\tilde{P}_{x_i}(\tau)\mathrm{d}\tau \\ - \int_{t_0}^{t} K_{x_i,x_i,+1}(t-\tau)\tilde{P}_{x_i}(\tau)\,\mathrm{d}\tau, \quad (4.18)\end{aligned}$$

which can be represented in a more compact form as

$$\partial_t|\tilde{P}(t)\rangle = \hat{\mathcal{L}}(t)|\tilde{P}(t)\rangle, \quad (4.19)$$

where $\hat{\mathcal{L}}(t)$ is a linear s-dependent integral operator and $|\tilde{P}(t)\rangle$ has components $\tilde{P}_{x_i}(t)$. The limit as $T\to\infty$ of $\ln\langle 1|\tilde{P}(t)\rangle/T$ (where $\langle 1|$ is a row vector with all entries equal to one) is the SCGF of J. Clearly, equation (4.19) does not conserve the product $\langle 1|\tilde{P}(t)\rangle$, except for $s=0$ when this corresponds to the condition $\sum_{x_i} P_{x_i}(t)=1$. The dynamics described by equation (4.18) is equivalent to the dynamics described by the GME (4.15), where the memory kernels corresponding to jumps that contribute a unit c in the total current are "biased", i.e., multiplied by a factor e^{-cs}. From linearity, it follows that the Laplace transformed kernels are

$$\mathrm{e}^{-cs}\overline{K}_{x_i,x_j,c}(\nu) = \mathrm{e}^{-cs}\overline{\psi}_{x_i,x_j,c}(\nu)\Big/\overline{\phi}_{x_j}(\nu)\ , \quad (4.20)$$

where $\overline{f}(\nu)=\int_0^\infty \mathrm{e}^{-\nu T} f(T)\mathrm{d}T$. This confirms that the modified dynamics can be simulated biasing directly the WTDs $\psi_{x_i,x_j,c}(t)$, i.e., multiplying them by e^{-cs}.

In section 1.4.1, we studied the behaviour of $I_{x_i}(t)$ and saw that, under some general conditions, it converges to zero in the limit as $t\to\infty$. We now consider the s-dependent case, and study the asymptotic behavior of $\tilde{I}_{x_i}(t)=\sum_{\mathsf{J}}\mathrm{e}^{-s\mathsf{J}} I_{(x_i,\mathsf{J})}(t)$. In the joint configuration–current space, the term encoding for the initial WTDs is, in the

Laplace space conjugated to time,

$$\begin{aligned}\overline{I}_{(x_i,J)}(\nu) = &\sum_{x_j\neq x_i} \overline{\psi'}_{x_i,x_j,0}(\nu)P_{(x_j,J)}(t_0) + \sum_{x_j,c=\pm1} \overline{\psi'}_{x_i,x_j,c}(\nu)P_{(x_j,J-c)}(t_0)\\ &- \left[\sum_{x_j\neq x_i} \overline{\psi'}_{x_j,x_i,0}(\nu) + \sum_{x_j,c=\pm1} \overline{\psi'}_{x_j,x_i,c}(\nu)\right] P_{(x_i,J)}(t_0)\\ &+ \sum_{x_j\neq x_i} \overline{\psi}_{x_j,x_i,0}(\nu)\frac{\overline{\phi'}_{x_i}(\nu)}{\overline{\phi}_{x_i}(\nu)}P_{(x_i,J)}(t_0) + \sum_{x_j,c=\pm1} \overline{\psi}_{x_j,x_i,c}(\nu)\frac{\overline{\phi'}_{x_i}(\nu)}{\overline{\phi}_{x_i}(\nu)}P_{(x_i,J-c)}(t_0)\\ &- \left[\sum_{x_j\neq x_i} \overline{\psi}_{x_i,x_j,0}(\nu)\frac{\overline{\phi'}_{x_j}(\nu)}{\overline{\phi}_{x_j}(\nu)} + \sum_{x_j,c=\pm1} \overline{\psi}_{x_i,x_j,c}(\nu)\frac{\overline{\phi'}_{x_j}(\nu)}{\overline{\phi}_{x_j}(\nu)}\right] P_{(x_j,J)}(t_0), \quad (4.21)\end{aligned}$$

hence,

$$\begin{aligned}\tilde{\overline{I}}_{x_i}(\nu) = &\left[\sum_{x_j\neq x_i} \overline{\psi'}_{x_i,x_j,0}(\nu) + \sum_{x_j,c=\pm1} \mathrm{e}^{-cs}\overline{\psi'}_{x_i,x_j,c}(\nu)\right] \tilde{P}_{x_j}(t_0)\\ &- \left[\sum_{x_j\neq x_i} \overline{\psi'}_{x_j,x_i,0}(\nu) + \sum_{x_j,c=\pm1} \overline{\psi'}_{x_j,x_i,c}(\nu)\right] \tilde{P}_{x_i}(t_0)\\ &+ \left[\sum_{x_j\neq x_i} \overline{\psi}_{x_j,x_i,0}(\nu)\frac{\overline{\phi'}_{x_i}(\nu)}{\overline{\phi}_{x_i}(\nu)} + \sum_{x_j,c=\pm1} \mathrm{e}^{-cs}\overline{\psi}_{x_j,x_i,c}(\nu)\frac{\overline{\phi'}_{x_i}(\nu)}{\overline{\phi}_{x_i}(\nu)}\right] \tilde{P}_{x_i}(t_0)\\ &- \left[\sum_{x_j\neq x_i} \overline{\psi}_{x_i,x_j,0}(\nu)\frac{\overline{\phi'}_{x_j}(\nu)}{\overline{\phi}_{x_j}(\nu)} + \sum_{x_j,c\pm1} \psi_{x_i,x_j,c}(\nu)\frac{\overline{\phi'}_{x_j}(\nu)}{\overline{\phi}_{x_j}(\nu)}\right] \tilde{P}_{x_j}(t_0). \quad (4.22)\end{aligned}$$

As at the beginning of the observation time the total current is zero, we can replace $\tilde{P}_{x_0}(t_0)$ with $P_{x_0}(t_0)$. Using WTDs that satisfy either (1.86) or (1.87), which were introduced in section 1.4, we again find that the limit as $\nu \to 0$ is finite, hence, according to the *final-value* theorem[2], $\tilde{I}_{x_i}(t)$ decays to zero in the long-time limit. However, the initial-condition term may still substantially affect the large deviation functionals and their numerical evaluation, as such a decay may be slow for certain choices of WTDs. In general, $\tilde{\overline{I}}_{x_i}(\nu)$ does *not* vanish even when $\overline{\psi'}_{x_i,x_j,c}(\nu) = \overline{\psi}_{x_i,x_j,c}(\nu)$. In fact, in this case we have

$$\begin{aligned}\tilde{\overline{I}}_{x_i}(\nu) = \sum_{x_i} \Big\{&(\mathrm{e}^{-s}-1)\left[\overline{\psi}_{x_j,x_i,+1}(\nu)P_{x_i}(t_0) + \overline{\psi}_{x_i,x_j,+1}(\nu)P_{x_j}(t_0)\right]\\ &+ (\mathrm{e}^{s}-1)\left[\overline{\psi}_{x_j,x_i,-1}(\nu)P_{x_i}(t_0) + \overline{\psi}_{x_i,x_j,-1}(\nu)P_{x_j}(t_0)\right]\Big\}, \quad (4.23)\end{aligned}$$

[2]This was also introduced in section 1.4.

which is in general non-zero (except for $s = 0$, when $\overline{I}_{x_i}(\nu) = 0$ is recovered). Consequently, the algorithm of section 4.2.2 must be iterated for sufficiently long time in order to neglect this finite-time contribution.

4.3.2 The Markovian case re-examined

The Markovian case is recovered for $K_{x_i,x_j,c}(t) = g_{x_i,x_j,c}\delta(t)$. Using such a kernel, equations (4.18) and (4.19) can be written as

$$\partial_t|\tilde{P}(t)\rangle = \tilde{\mathbf{G}}|\tilde{P}(t)\rangle, \tag{4.24}$$

where $\tilde{\mathbf{G}}$ is the s-modified stochastic generator of the Markov process with time-independent rates g_{x_i,x_j} and components[3]

$$\left[\tilde{\mathbf{G}}\right]_{x_i,x_j} = g_{x_i,x_j,0} + \mathrm{e}^{-s}g_{x_i,x_j,+1} + \mathrm{e}^{s}g_{x_i,x_j,-1}, \tag{4.25}$$

$$\left[\tilde{\mathbf{G}}\right]_{x_i,x_i} = \mathrm{e}^{-s}g_{x_i,x_i,+1} + \mathrm{e}^{s}g_{x_i,x_i,-1} - g_{x_i,x_i,-1} - g_{x_i,x_i,+1} - g_{x_i}, \tag{4.26}$$

where $g_{x_i} = \sum_{c,x_j\neq x_i} g_{x_j,x_i,c}$ is the rate of escape from x_i. This shows that biasing the rates is consistent with biasing the WTDs (see also some related discussions in [17]). However, from a numerical point of view, the latter choice remains convenient even for the Markovian case, as it avoids us having to define the modified transition probabilities of Lecomte and Tailleur [29]. To see this, we consider the biased Markovian WTD

$$\tilde{\psi}_{x_i,x_j,c}(\tau) = \mathrm{e}^{-cs}g_{x_i,x_j,c}\exp\left(-g_{x_j}\tau\right), \tag{4.27}$$

which is the product of an exponential probability density $\psi_{x_j}(\tau) = g_{x_j}\exp\left(-g_{x_j}\tau\right)$, a time-independent probability mass $p_{x_i,x_j,c} = g_{x_i,x_j,c}/g_{x_j}$, and a simple cloning factor e^{-cs}. These specify the two steps of the standard Doob–Gillespie algorithm for Markov processes (see section 1.7.1), followed by a cloning step of weight e^{-cs}. Another legitimate choice is to define the biased rates $\tilde{g}_{x_i,x_j,c} = \mathrm{e}^{-cs}g_{x_i,x_j,c}$ and $\tilde{g}_{x_j} = \sum_{x_i,c}\tilde{g}_{x_i,x_j,c}$ to write

$$\tilde{\psi}_{x_i,x_j,c}(\tau) = \exp\left[\tau\left(\tilde{g}_{x_j} - g_{x_j}\right)\right]\tilde{g}_{x_i,x_j,c}\exp\left(-\tilde{g}_{x_j}\tau\right). \tag{4.28}$$

With such an arrangement, we recognise the scheme of Lecomte and Tailleur [29], i.e., at each temporal step, the configuration evolves according to a stochastic generator with rates $\tilde{g}_{x_i,x_j,c}$, (this is equivalent to the steps 2) and 3) of the algorithm in section 1.7.2)

[3] Note that the equation (4.26), and also the earlier (4.18), takes into account events that do not alter the configuration, but modify the current statistics. This slightly generalises the treatment of section 1.7.2, where the change in current is always accompanied by a configuration change.

and the ensemble is modified with the cloning factor $\exp\left[\tau\left(\tilde{g}_{x_j} - g_{x_j}\right)\right]$, as in the cloning step 4) of section 1.7.2. As the cloning factor here is exponential in time, during long intervals the relative number of new clones can be large. This can cause major finite-ensemble errors, which are shown to be important, e.g., in reference [11, 26]. Conversely, an implementation based on equation (4.27) seems to be one way to reduce (but not completely eliminate) such a problem.

Noticeably, for exponentially distributed waiting times we have

$$\overline{\psi}_{x_i,x_j,c}(\nu) = \frac{\beta_{x_i,x_j,c}}{\beta_{x_j} + \nu}, \tag{4.29}$$

$$\overline{\psi'}_{x_i,x_j,c}(\nu) = \frac{\beta'_{x_i,x_j,c}}{\beta'_{x_j} + \nu}. \tag{4.30}$$

Substituting into equation (4.23), it is straightforward to show that the finite-time effects due to $I_{x_i}(t)$ are only minor. In fact, the resulting exact equation

$$\begin{aligned}\tilde{\tilde{I}}_{x_i}(\nu) = &\sum_{x_j \neq x_i} \frac{\tilde{\beta}_{x_j,x_i,0} - \tilde{\beta'}_{x_j,x_i,0}}{\tilde{\beta'}_{x_i} + \nu} P_{x_i}(t_0) + \sum_{x_j,c=\pm 1} \frac{\tilde{\beta}_{x_j,x_i,c} - \tilde{\beta'}_{x_j,x_i,c}}{\tilde{\beta'}_{x_i} + \nu} P_{x_i}(t_0) \\ &+ \sum_{x_j \neq x_i} \frac{\tilde{\beta'}_{x_i,x_j,0} - \tilde{\beta}_{x_i,x_j,0}}{\tilde{\beta'}_{x_j} + \nu} P_{x_j}(t_0) + \sum_{x_j,c=\pm 1} \frac{\tilde{\beta'}_{x_i,x_j,c} - \tilde{\beta}_{x_i,x_j,c}}{\tilde{\beta'}_{x_j} + \nu} P_{x_j}(t_0),\end{aligned} \tag{4.31}$$

is sum of Laplace transforms of exponential functions and implies an exponential decay of $\tilde{I}_{x_i}(t)$ to zero.

4.3.3 SCGF as pole of the partition function

We report here a neat procedure to find the long-time behaviour of current fluctuations in semi-Markov processes proposed in Andrieux and Gaspard [2]. This will be used to test the exactness of the non-Markovian cloning method. In this section, we derive such a procedure with the minor extension to semi-Markov processes with special (primed) initial WTDs. The starting point is the explicit form for the biased probability of the

configuration x after a time $t-t_0$ (since the reference instant t_0),

$$\begin{aligned}\tilde{P}_x(t) &= \sum_{x_0}\delta_{x,x_0}\phi'_{x_0}(t-t_0)P_{x_0}(t_0) + \int_{t_0}^{t}\mathrm{d}t_1\sum_{x_0,x_1}\delta_{x,x_1}\phi_{x_1}(t-t_1)\tilde{\psi}'_{x_1,x_0}(t_1-t_0)P_{x_0}(t_0)\\ &+\int_{t_0}^{t}\mathrm{d}t_1\int_{t_1}^{t}\mathrm{d}t_2\sum_{x_0,x_1,x_2}\delta_{x,x_2}\phi_{x_2}(t-t_2)\tilde{\psi}_{x_2,x_1}(t_2-t_1)\tilde{\psi}'_{x_1,x_0}(t_1-t_0)P_{x_0}(t_0)\\ &+\int_{t_0}^{t}\mathrm{d}t_1\int_{t_1}^{t}\mathrm{d}t_2\int_{t_2}^{t}\mathrm{d}t_3\sum_{x_0,x_1,x_2,x_3}\delta_{x,x_3}\phi_{x_3}(t-t_3)\tilde{\psi}_{x_3,x_2}(t_3-t_2)\tilde{\psi}_{x_2,x_1}(t_2-t_1)\tilde{\psi}'_{x_1,x_0}(t_1-t_0)P_{x_0}(t_0)\\ &+\quad\ldots,\end{aligned}\tag{4.32}$$

which gives the partition function $Z(s,t)=\sum_x\tilde{P}_x(t)$. Equation (4.32) can be Laplace transformed recursively to yield

$$\begin{aligned}\tilde{\overline{P}}_x(\nu) &= \sum_{x_0}\delta_{x,x_0}\overline{\phi}'_{x_0}(\nu)P_{x_0}(t_0)+\sum_{x_0,x_1}\delta_{x,x_1}\overline{\phi}_{x_1}(\nu)\tilde{\overline{\psi}}'_{x_1,x_0}(\nu)P_{x_0}(t_0)\\ &+\sum_{x_0,x_1,x_2}\delta_{x,x_2}\overline{\phi}_{x_2}(\nu)\tilde{\overline{\psi}}_{x_2,x_1}(\nu)\tilde{\overline{\psi}}'_{x_1,x_0}(\nu)P_{x_0}(t_0)\\ &+\sum_{x_0,x_1,x_2,x_3}\delta_{x,x_3}\overline{\phi}_{x_3}(\nu)\tilde{\overline{\psi}}_{x_3,x_2}(\nu)\tilde{\overline{\psi}}_{x_2,x_1}(\nu)\tilde{\overline{\psi}}'_{x_1,x_0}(\nu)P_{x_0}(t_0)\\ &+\quad\ldots,\end{aligned}\tag{4.33}$$

which can be compactly written as

$$\begin{aligned}\overline{P}_x(\nu) &= \sum_{x_0}\delta_{x,x_0}\overline{\phi}'_{x_0}(\nu)P_{x_0}(t_0)\\ &+\sum_{x_0}\sum_{n=1}^{\infty}\sum_{x_1,\ldots,x_n}\delta_{x,x_n}\overline{\phi}_{x_n}(\nu)\tilde{\overline{\psi}}_{x_n,x_{n-1}}(\nu)\ldots\tilde{\overline{\psi}}_{x_2,x_1}(\nu)\tilde{\overline{\psi}}'_{x_1,x_0}(\nu)P_{x_0}(t_0).\end{aligned}\tag{4.34}$$

Introducing the matrices $\overline{\boldsymbol{\phi}}(\nu)$, $\tilde{\overline{\boldsymbol{\psi}}}(\nu)$, $\overline{\boldsymbol{\phi}}'(\nu)$, $\tilde{\overline{\boldsymbol{\psi}}}'(\nu)$, with entries

$$\left[\overline{\boldsymbol{\phi}}(\nu)\right]_{x_j,x_i}=\delta_{x_j,x_i}\overline{\phi}_{x_i}(\nu),\tag{4.35}$$

$$\left[\overline{\boldsymbol{\psi}}(\nu)\right]_{x_j,x_i}=\tilde{\overline{\psi}}_{x_j,x_i}(\nu),\tag{4.36}$$

$$\left[\overline{\boldsymbol{\phi}}'(\nu)\right]_{x_j,x_i}=\delta_{x_j,x_i}\overline{\phi}'_{x_i}(\nu),\tag{4.37}$$

$$\left[\overline{\boldsymbol{\psi}}'(\nu)\right]_{x_j,x_i}=\tilde{\overline{\psi}}'_{x_j,x_i}(\nu),\tag{4.38}$$

respectively, equation (4.34) can be written as

$$\overline{P}_x(\nu) = \sum_{x_0} \left[\overline{\phi'}(\nu)\right]_{x,x_0} P_{x_0}(t_0) + \sum_{x_0} \left[\sum_{n=1}^{\infty} \overline{\phi}(\nu)\tilde{\overline{\psi}}(\nu)^{n-1}\tilde{\overline{\psi'}}(\nu)\right]_{x,x_0} P_{x_0}(t_0) \tag{4.39}$$

$$= \sum_{x_0} \left[\overline{\phi'}(\nu) + \sum_{n=1}^{\infty} \overline{\phi}(\nu)\tilde{\overline{\psi}}(\nu)^{n-1}\tilde{\overline{\psi'}}(\nu)\right]_{x,x_0} P_{x_0}(t_0) \tag{4.40}$$

$$= \sum_{x_0} \left[\overline{\phi'}(\nu) + \overline{\phi}(\nu)\frac{1}{\mathbb{1} - \tilde{\overline{\psi}}(\nu)}\tilde{\overline{\psi'}}(\nu)\right]_{x,x_0} P_{x_0}(t_0). \tag{4.41}$$

In order to obtain the dominant mode in the long-time limit, one can study the poles of the matrix of equation (4.41) with entries $\left[\overline{\phi'}(\nu) + \overline{\phi}(\nu)\tilde{\overline{\psi'}}(\nu)/(\mathbb{1} - \tilde{\overline{\psi}}(\nu))\right]_{x,x_0}$. Such poles occur where the determinant of $\mathbb{1} - \overline{\psi}(\nu)$ is zero. The main limitation of this method is that it only applies to semi-Markov processes on finite configuration spaces.

4.3.4 Discrete-time case

We now consider discrete-time processes. A discrete-time chain can be seen as a stochastic process in continuous time where the next jump occurs after a constant waiting time of one unit. Such a scenario can be represented by means of a semi-Markov process with WTDs $\psi_{x_n,x_{n-1}}[\tau; w(t)] = p_{x_n,x_{n-1}}[w(t)]\psi_{x_{n-1}}(\tau)$, where $p_{x_n,x_{n+1}}[w(t)]$ is an entry of a transition matrix and $\psi_{x_{n-1}}(\tau) = \delta(\tau - 1)$ is the Dirac delta measure translated by 1. In fact, the procedure of section 4.2.2 can be implemented with reasonable accuracy by setting, e.g.,

$$\psi_{x_{n-1}}(\tau) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-(\tau - 1)^2/(2\sigma^2)\right], \tag{4.42}$$

with $\sigma \ll 1$. In particular, a discrete-time *Markov* chain can be seen as a special DTI semi-Markov process, since the transition probabilities do not depend on $w(t)$. However, such a continuous-time implementation neglects the major computational advantage of dealing with discrete time, namely, all the ensemble elements can be updated simultaneously. This feature can be used to prevent a single clone replacing a macroscopic fraction of the ensemble, thus reducing finite size effects. Therefore, we suggest the following parallel algorithm:

1) Set up an ensemble of N clones and initialise each to its own random configuration x_0. Also, initialise a unique counter to $n = 1$, the variable C to zero, and each element of an array $\mathbf{C}$ of length N to 1.

2) For each clone, update the configuration from x_{n-1} to x_n according to the mass

$p_{x_n,x_{n-1}}[1; w(n-1)]$. Store the individual values of $\mathrm{e}^{-s\theta_{x_n,x_{n-1}}}$ in $\mathbf{C}$.

3) **Cloning step.** Compute the arithmetic mean y of all the entries of $\mathbf{C}$. Perform a weighted random sampling with repetition (see, e.g., reference [15]) of N clones from the ensemble, according to their weights $\mathbf{C}$. This sample replaces the existing ensemble.

4) Increment C to $C + \ln(y)$. Update n to $n+1$ and reiterate from 2, until n reaches the desired simulation time.

The SCGF is recovered as $-C/n$ for large n. As the sampling at step 3) is performed simultaneously for all the clones, it is very unlikely for a single clone to replace all the remaining ones, even in the presence of a strong bias. We argue that this further reduces the finite ensemble effects.

In continuous time an equivalent strategy is to mimic the discrete-time steps, as in, e.g., reference [5], so each trajectory evolves independently for a constant interval Δt; in this case, the product of the cloning factors encountered during the interval, as well as the time elapsed since the last jump must be stored. This permits the application of the cloning step to all clones simultaneously.

Finally, it is worth making the link to the procedure originally proposed by C. Giardinà, J. Kurchan, and L. Peliti [18]. In their seminal paper, the use of the cloning method to probe large deviation functionals for Markov processes is proposed and the general idea is implemented in discrete-time. Indeed, for the Markovian case, we can arrange the biased WTD as

$$\tilde{\psi}_{x_i,x_j,c}(\tau) = \frac{\sum_{x_k,c'} \tilde{g}_{x_k,x_j,c'}}{\sum_{x_k,c'} g_{x_k,x_j,c'}} \frac{\tilde{g}_{x_i,x_j,c}}{\sum_{x_k,c'} \tilde{g}_{x_k,x_j,c'}} \delta(\tau - 1). \tag{4.43}$$

This suggests the following steps for each ensemble element: increase the time by one unit, change the state according to the modified transition probability $\tilde{g}_{x_i,x_j,c} / \sum_{x_k,c'} \tilde{g}_{x_k,x_j,c'}$ and modify the ensemble population according to a cloning factor

$$\sum_{x_k,c'} \tilde{g}_{x_k,x_j,c'} \Big/ \sum_{x_k,c'} \beta_{x_k,x_j,c'} \ , \tag{4.44}$$

which only depends on the departure configuration x_j, as indeed explained in reference [18].

4.4 Test on non-Markovian toy models

We now test the non-Markovian cloning procedure against three non-Markovian models, whose exact large deviations are known from the literature or can be deduced from Markovian models.

4.4.1 Semi-Markov models for ion-channel gating with and without DTI

The current through an ion channel in a cellular membrane can be modelled with only two states, corresponding to the gate being singly occupied (x_1) or empty (x_0); an ion can enter or leave this channel via the left (L) or right (R) boundary and non-exponential waiting times lead to a complex behaviour [3]. Specifically, we denote the WTD for a particle that succeeds in entering (or leaving) through the boundary L by $\psi_{x_1,x_0,1}(\tau)$ (or $\psi_{x_0,x_1,-1}(\tau)$) with respective density $\psi_{x_1,x_0,0}(\tau)$ (or $\psi_{x_0,x_1,0}(\tau)$) for the boundary R. The rightwards current is measured by a counter that increases (decreases) by one when a particle enters (leaves) the system through the boundary L. Its exact SCGF is obtained numerically by D. Andrieux and P. Gaspard [2] as the leading pole of the time-Laplace transform of $Z(s,t)$ (procedure detailed in section 4.3.3), for the DTI-case $\psi_{x_i,x_j,c}(\tau) = p_{x_i,x_j,c}\psi_{x_j}(\tau)$ with $\sum_{c=-1,0} p_{x_0,x_1,c} = \sum_{c=0,1} p_{x_1,x_0,c} = 1$, and the particular choice $\psi_{x_j}(\tau) = g(\tau; k_j, \lambda_j)$, where

$$g(\tau; k, \lambda) = \frac{\lambda^k \tau^{k-1}}{\Gamma(k)} \exp(-\lambda\tau) \tag{4.45}$$

is the PDF of a Gamma distribution with shape k and scale $1/\lambda$ and

$$\Gamma(k) = \int_0^\infty x^{k-1} \mathrm{e}^x \, \mathrm{d}x$$

is the Gamma function. The Markovian case is recovered for $k = 1$.

We intend to find the points where the determinant of $\mathbb{1} - \overline{\boldsymbol{\psi}}(\nu)$ is zero, as detailed in section 4.3.3 and reference [2]. The Laplace transforms of the Gamma WTDs are

$$\overline{\psi}_{x_j}(\nu) = \left(\frac{\lambda_j}{\lambda_j + \nu}\right)^{k_j}. \qquad \text{for } i = 1, 2 \tag{4.46}$$

Notice that if the shape parameter k_j is integer, then $\overline{\psi}_{x_j}(\nu)$ has k_j real poles and an Erlang distribution with k_j stages is obtained. The total Laplace-transformed biased

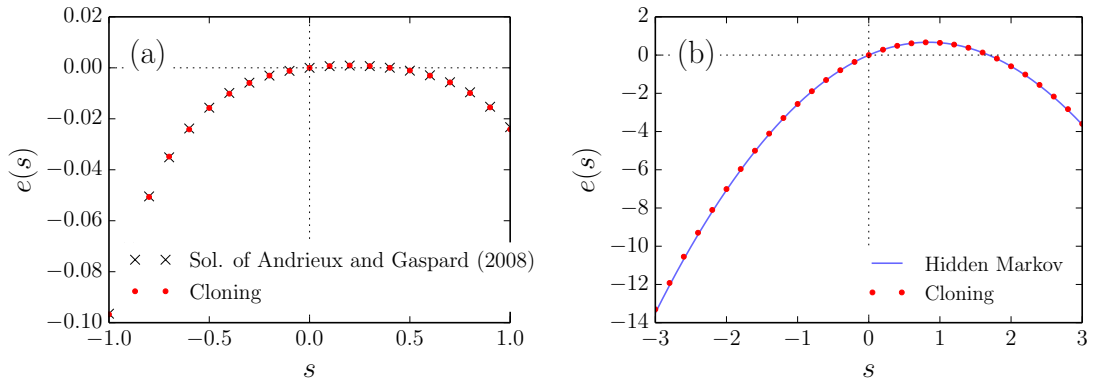

Figure 4.2: SCGF of current in ion channel. (a) DTI model with $(k_0, \lambda_0, k_1, \lambda_1) = (0.1, 0.01, 1, 1)$ and $(p_{x_1,x_0,1}, p_{x_0,x_1,0}, p_{x_0,x_1,-1}, p_{x_1,x_0,0}) = (0.5, 0.6, 0.4, 0.5)$; the cloning result is consistent with the solution given in [2]. (b) non-DTI model with Markov representation and inverse scales $(\lambda_0^L, \lambda_0^R, \lambda_1^L, \lambda_1^R) = (20, 10, 10, 20/3)$; The cloning reproduces the leading eigenvalue of the Markovian s-modified generator. In both cases $N = 10^3$ and $t = 10^3$.

WTDs are;

$$\tilde{\overline{\psi}}_{x_0,x_1}(\nu) = (p_{x_0,x_1,0} + p_{x_0,x_1,1}\mathrm{e}^{-s})\overline{\psi}_{x_1}(\nu), \tag{4.47}$$

$$\tilde{\overline{\psi}}_{x_1,x_0}(\nu) = (p_{x_1,x_0,0} + p_{x_1,x_0,-1}\mathrm{e}^{s})\overline{\psi}_{x_0}(\nu); \tag{4.48}$$

these are the off-diagonal elements of the matrix $\mathbb{1} - \overline{\boldsymbol{\psi}}(\nu)$, whose determinant $1 - \tilde{\overline{\psi}}_{x_1,x_0}(\nu)\tilde{\overline{\psi}}_{x_0,x_1}(\nu)$ is zero when

$$\left(\frac{\lambda_1 + \nu}{\lambda_1}\right)^{k_1}\left(\frac{\lambda_0 + \nu}{\lambda_0}\right)^{k_0} = (p_{x_0,x_1,0} + p_{x_0,x_1,1}\mathrm{e}^{-s})(p_{x_0,x_1,0} + p_{x_0,x_1,1}\mathrm{e}^{-s}). \tag{4.49}$$

The leading zero of equation (4.49), computed using the Newton–Raphson method, is plotted in figure 4.2(a) along with the cloning results. Notably, the cloning method of section 4.2 can be implemented for any WTD, as only a bias of e^{s} for ions leaving the channel leftwards and a bias e^{-s} for ions entering from left are needed. The figure shows that the cloning reproduces, within numerical accuracy, the solution in ν of equation (4.49).

In a quest for more general classes of models to illustrate the power of our approach, we now relax the constraint of DTI and assume that each transition can be triggered independently by two mechanisms, corresponding to the two boundaries. We still assume that memory of the previous history is lost as soon as the system changes state, thus preserving the semi-Markov nature. At the instant when the gate is emptied, a particle attempts to enter the system from the left boundary after a waiting time T_0^L with density distribution $\Psi_0^L(\tau)$, while another particle attempts to arrive from the right boundary

after a time T_0^R distributed according to $\Psi_0^R(\tau)$. The waiting times T_1^L and T_1^R, as well as the densities $\Psi_1^L(\tau)$ and $\Psi_1^R(\tau)$ are defined similarly. In order to have a right (left) jump during the interval $[\tau, \tau + \mathrm{d}\tau)$, we also require that the left (right) mechanism remains silent until time τ. Consequently, the WTDs are

$$\psi_{x_1,x_0,0}(\tau) = \Psi_0^R(\tau)\Phi_0^L(\tau), \tag{4.50}$$
$$\psi_{x_1,x_0,-1}(\tau) = \Psi_0^L(\tau)\Phi_0^R(\tau), \tag{4.51}$$
$$\psi_{x_0,x_1,0}(\tau) = \Psi_1^R(\tau)\Phi_1^L(\tau), \tag{4.52}$$
$$\psi_{x_0,x_1,1}(\tau) = \Psi_1^L(\tau)\Phi_1^R(\tau), \tag{4.53}$$

where $\Phi_j^{(\rho)}(\tau) = \int_\tau^\infty \Psi_j^{(\rho)}(t)\,\mathrm{d}t$ are survival probabilities, with ρ denoting the mechanism L or R. As a concrete choice, we again assign a Gamma probability distribution to the waiting time of each event,

$$\Psi_j^{(\rho)}(\tau) = g(\tau; k_j^{(\rho)}, \lambda_j^{(\rho)}), \tag{4.54}$$

so that the survival probabilities are

$$\Phi_j^{(\rho)}(\tau) = \Gamma(k_j^{(\rho)}, \lambda_j^{(\rho)}\tau)/\Gamma(k_j^{(\rho)}), \tag{4.55}$$

where $\Gamma(k, x)$ is the upper incomplete Gamma function. The time to the next jump, given that the system just reached state x_j (i.e., its age is zero) is $\min\{T_j^L, T_j^R\}$ and is associated with the total survival probability,

$$\phi_{x_j}(\tau) = \Phi_j^L(\tau)\Phi_j^R(\tau). \tag{4.56}$$

Once the transition time is known, either the left or right trigger is chosen, according to the age-dependent rates

$$\beta_j^{(\rho)}(\tau) = g(\tau; k_j^{(\rho)}, \lambda_j^{(\rho)})\Gamma(k_j^{(\rho)})/\gamma(k_j^{(\rho)}, \lambda_j^{(\rho)}\tau), \tag{4.57}$$

where $\gamma(k, x)$ is the lower incomplete Gamma function. The SCGF of the left current is computed by biasing the WTDs $\psi_{x_1,x_0,1}(\tau)$ and $\psi_{x_0,x_1,-1}(\tau)$ with $\mathrm{e}^{\mp s}$, respectively.

While the implementation of the method of section 4.2.2 remains straightforward for this model, a general solution for the exact SCGF is missing. We thus specialise to the case with integer shapes $k_0^R = k_1^R = k_0^L = k_1^L = 2$, where the Laplace transform of $\psi_{x_i,x_j}(\tau) = \sum_c \psi_{x_i,x_j,c}(\tau)$ is a rational function of ν, viz.,

$$\overline{\psi}_{x_j,x_i}(\nu) = \frac{(\lambda_i^L)^2(\nu + 3\lambda_i^R + \lambda_i^L)}{\left(\nu + \lambda_i^R + \lambda_i^L\right)^3} + \frac{(\lambda_i^R)^2(\nu + \lambda_i^R + 3\lambda_i^L)}{\left(\nu + \lambda_i^R + \lambda_i^L\right)^3}; \tag{4.58}$$

its first term corresponds to the right boundary, while the second one corresponds to the left boundary[4]. Notice that there is no dependence on the arrival state x_j, the model being defined on a two-state configuration space. Equation (4.58) can be conveniently written as

$$\overline{\psi}_{x_j,x_i}(\nu) = \alpha_{i,2}\frac{(\lambda_i^R+\lambda_i^L)^2}{(\nu+\lambda_i^R+\lambda_i^L)^2} + \alpha_{i,3}\frac{(\lambda_i^R+\lambda_i^L)^3}{(\nu+\lambda_i^R+\lambda_i^L)^3}, \tag{4.59}$$

with

$$\alpha_{i,2} = \frac{(\lambda_i^R)^2+(\lambda_i^L)^2}{\left(\lambda_i^R+\lambda_i^L\right)^2}, \tag{4.60}$$

$$\alpha_{i,3} = \frac{2\lambda_i^R\lambda_i^L}{\left(\lambda_i^R+\lambda_i^L\right)^2}, \tag{4.61}$$

which clearly defines a Coxian distribution of the type seen in equation (1.37) (notice that $\alpha_{i,2}+\alpha_{i,3}=1$). To study the effect of boundaries we separately decompose in partial fractions the left and right WTD contributions of equation (4.58), i.e.,

$$\overline{\psi}_{x_j,x_i}(\nu) = \frac{\left(\lambda_i^L\right)^2 2\lambda^R}{\left(\nu+\lambda_i^R+\lambda_i^L\right)^3} + \frac{\left(\lambda_i^L\right)^2}{\left(\nu+\lambda_i^R+\lambda_i^L\right)^2} + \frac{\left(\lambda_i^R\right)^2 2\lambda^L}{\left(\nu+\lambda_i^R+\lambda_i^L\right)^3} + \frac{\left(\lambda_i^R\right)^2}{\left(\nu+\lambda_i^R+\lambda_i^L\right)^2}, \tag{4.62}$$

which can be rearranged as

$$\begin{aligned}\overline{\psi}_{x_j,x_i}(\nu) =& \alpha_{i,2}^L\frac{\left(\lambda_i^R+\lambda_i^L\right)^2}{\left(\nu+\lambda_i^R+\lambda_i^L\right)^2} + \alpha_{i,3}^L\frac{\left(\lambda_i^R+\lambda_i^L\right)^3}{\left(\nu+\lambda_i^R+\lambda_i^L\right)^3} \\ &+ \alpha_{i,2}^R\frac{\left(\lambda_i^R+\lambda_i^L\right)^2}{\left(\nu+\lambda_i^R+\lambda_i^L\right)^2} + \alpha_{i,3}^R\frac{\left(\lambda_i^R+\lambda_i^L\right)^3}{\left(\nu+\lambda_i^R+\lambda_i^L\right)^3},\end{aligned} \tag{4.63}$$

where

$$\alpha_{i,2}^L = \frac{\left(\lambda_i^L\right)^2}{\left(\lambda_i^R+\lambda_i^L\right)^2}, \tag{4.64}$$

$$\alpha_{i,3}^L = \frac{2\left(\lambda_i^L\right)^2\lambda_i^R}{\left(\lambda_i^R+\lambda_i^L\right)^3}, \tag{4.65}$$

$$\alpha_{i,2}^R = \frac{\left(\lambda_i^R\right)^2}{\left(\lambda_i^R+\lambda_i^L\right)^2}, \tag{4.66}$$

$$\alpha_{i,3}^R = \frac{2\left(\lambda_i^R\right)^2\lambda_i^L}{\left(\lambda_i^R+\lambda_i^L\right)^3}. \tag{4.67}$$

Notice that $\alpha_{i,2}^L+\alpha_{i,3}^L+\alpha_{i,2}^R+\alpha_{i,3}^R=1$. The first and second terms correspond to left jumps,

[4]This can be verified by Laplace-transforming the two terms in $\sum_c \psi_{x_i,x_j,c}(\tau)$ separately.

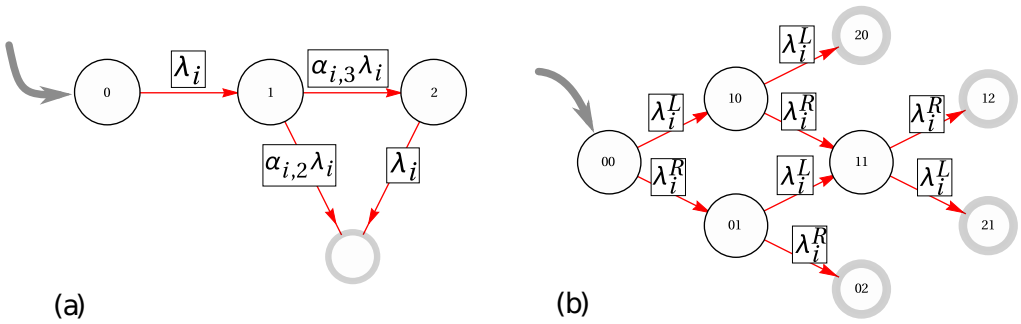

Figure 4.3: Two graphical representations of the WTD (4.59). The waiting time is equal to the adsorption time of a random walker from the leftmost site to any of the grey sites.

while the third and fourth terms correspond to right jumps. We also underline that the choice (4.64)–(4.67) is only one of the possible decompositions of the WTD (4.58). It follows, straightforwardly, that

$$\overline{\psi}_{x_i,x_j}(\nu) = \left(\alpha_{j,2}^R + \alpha_{j,2}^L\right)\left(\frac{\lambda_j}{\nu+\lambda_j}\right)^2 + \left(\alpha_{j,3}^R + \alpha_{j,3}^L\right)\left(\frac{\lambda_j}{\nu+\lambda_j}\right)^3, \tag{4.68}$$

where we used, for convenience, λ_j instead of $\lambda_j^L + \lambda_j^R$. A comparison with (1.37) shows that it corresponds to the case with three stages (i.e., $k = 3$), $p_0 = p_1 = 0$, and $p_2 = \alpha_{i,2}^R + \alpha_{i,2}^L$. Hence, the jump from x_i to x_j can be modelled as a process of three stages, in each of which the system is trapped for an exponentially distributed time with rate λ_i, as in explained in section 1.3. At time zero, with probability 1, the system enters the first stage and waits there. Then, again with probability 1, it enters a second identical stage. After leaving the second stage, the escape occurs immediately with probability p_2, or the system enters the third and last phase with probability $1 - p_2$. Hence the WTD is the time to absorption of the Markov process with the transition graph of figure 4.3(a), given that we start at state 0.

Recalling the notion of trigger, it is possible to build an alternative but equivalent absorbing Markov process with the same time to absorption. We think of each of the two Gamma triggers (R or L) as a device with two exponential stages (with rate λ_i^R or λ_i^L). The escape occurs when either of the two triggers leaves the last stage. The transition graph of the associated Markov process is shown in figure 4.3(b). Hence, the two-state semi-Markov process with WTD (4.50) and (4.53) can be seen as a six-state Markov process. With the phase-type representations of $\psi_{x_1,x_0}(\tau)$ and $\psi_{x_0,x_1}(\tau)$, it is straightforward to build a Markov transition graph of the full model. In order to study the non-equilibrium aspects, we need to distinguish the contributions of the two

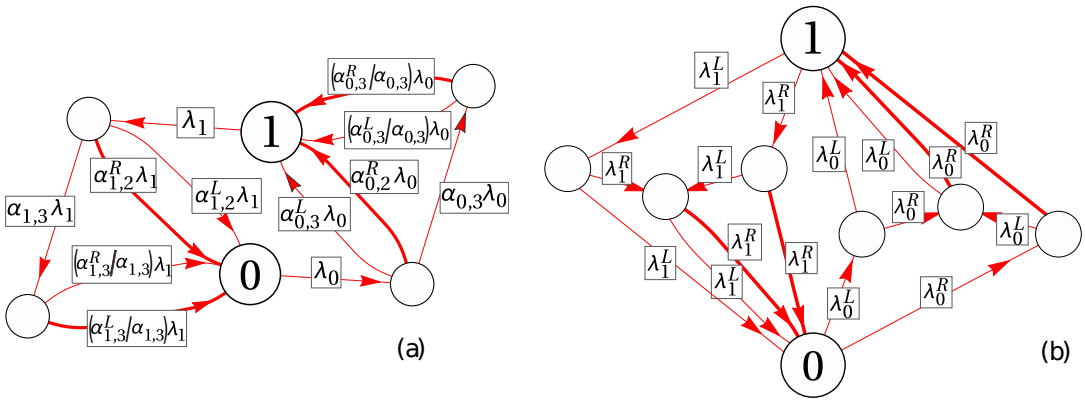

Figure 4.4: Graphical representations of the non-DTI ion-channel model with hidden states. The bonds corresponding to biased rates are drawn in thick lines. The modified generators associated with these two models have the same leading eigenvalue.

boundaries L and R. Hence, in order to obtain the s-modified generator and find the SCGF of figure 4.2(b), we only bias the true/visible transitions of type L. The resulting Markov representations are then encoded in the multi-graphs of figure 4.4.

The linearity of the Laplace transform permits the distinction of the left and right contributions in (4.68), hence it remains easy to bias the rates that correspond to a change in J and the SCGF can be exactly found as the leading eigenvalue of an s-modified Markovian stochastic generator. Figure 4.2(b) shows convincing agreement of our cloning method with this exact approach.

4.4.2 TASEP with history dependence

More general non-Markovian systems are those whose WTDs depend on events occurred during the whole observation time. Systems in this class are the "elephant" random walk [38] and its analogues [24, 36, 42], where the transition probabilities at time t depend on the history through the time-averaged current $j(t)$. We focus here on an IPS with such current-dependent rates, namely the TASEP of Harris [23].

Non-Markovian interacting particle systems can be described by assigning a trigger for attempts with WTD $\Psi_i[\tau; w(t)]$ and a corresponding survival function $\Phi_i[\tau; w(t)]$ to each elementary event i that controls the particle dynamics. The probability density that the next transition is of type i and occurs in the time interval $[t+\tau, t+\tau+dt)$, given

that, for each j, a time τ_j has elapsed since the last event of type j, is given by[5]

$$\psi_i[\tau; w(t)|\tau_0, \tau_1, \tau_2, \ldots] = \Psi_i[\tau + \tau_i; w(t)|\tau_i] \prod_{j\neq i, j=1,2,\ldots} \Phi_j[\tau + \tau_j; w(t)|\tau_j], \tag{4.69}$$

where $\Psi_i[\tau + \tau_i; w(t)|\tau_i] = \Psi_i[\tau + \tau_i; w(t)]/\Phi_i[\tau_i; w(t)]$ and $\Phi_i[\tau + \tau_i; w(t)|\tau_i] = \Phi_i[\tau + \tau_i; w(t)]/\Phi_i[\tau_i; w(t)]$. With exact expressions for these WTDs, we can implement the algorithms of section 4.2.

As mentioned in section 2.1, the TASEP consists of a one-dimensional lattice of length L, where each lattice site l, $1 \leq l \leq L$, can be either empty ($\eta_l = 0$) or occupied by a particle ($\eta_l = 1$). We assume that particles on a site $l < L$ are driven rightwards. In a non-Markovian variant of the model, they attempt a bulk jump to site $l + 1$ with WTD $\Psi_\mathrm{b}[\tau; w(t)]$, the attempt being successful if $\eta_{l+1} = 0$, as in [13, 21, 27]. With open boundaries, a particle that reaches the rightmost site L leaves the system with WTD $\Psi_L[\tau; w(t)]$. Also, as soon as $\eta_1 = 0$, a further boundary mechanism turns on and particles arrive on the leftmost site with WTD $\Psi_0[\tau; w(t)]$. The special choice $\Psi_0[\tau; w(t)] = \alpha \mathrm{e}^{-\alpha\tau}$, $\Psi_\mathrm{b}[\tau; w(t)] = p\mathrm{e}^{-p\tau}$, and $\Psi_L[\tau; w(t)] = \beta \mathrm{e}^{-\beta\tau}$ corresponds to the standard Markovian TASEP with constant left, bulk and right rates α, p, and β.

Let us provide more details about this model. From equation (4.69) we get

$$\psi_i[\tau; w(t)|\tau_0, \tau_1, \tau_2, \ldots] = \frac{\Psi_i[\tau + \tau_i; w(t)]}{\Phi_i[\tau_i; w(t)]} \prod_{j\neq i, j=0,1,2,\ldots} \frac{\Phi_j[\tau + \tau_j; w(t)]}{\Phi_j[\tau_j; w(t)]}. \tag{4.70}$$

which also yield, multiplying and dividing the r.h.s. by $\Phi_i[\tau_i + \tau; w(t)]$,

$$\psi_i[\tau; w(t)|\tau_0, \tau_1, \tau_2, \ldots] = \frac{\Psi_i[\tau + \tau_i; w(t)]}{\Phi_i[\tau_i + \tau; w(t)]} \prod_{j=0,1,2,\ldots} \frac{\Phi_j[\tau + \tau_j; w(t)]}{\Phi_j[\tau_j; w(t)]}. \tag{4.71}$$

In equation (4.71) we recognise the total survival probability

$$\prod_{j=0,1,2,\ldots} \frac{\Phi_j[\tau + \tau_j; w(t)]}{\Phi_j[\tau_j; w(t)]} \tag{4.72}$$

and the age-specific hazard

$$\frac{\Psi_i[\tau + \tau_i; w(t)]}{\Phi_i[\tau + \tau_i; w(t)]}. \tag{4.73}$$

[5] In a slight abuse of notation, we continue to use the full history $w(t)$ as a parameter but explicitly show the conditioning on the τ_is.

By dividing and multiplying the r.h.s. of equation (4.71) by the sum

$$\sum_i \frac{\Psi_i[\tau+\tau_i;w(t)]}{\Phi_i[\tau+\tau_i;w(t)]}, \tag{4.74}$$

we obtain the convenient form (4.11), i.e., the product of an inter-event time distribution

$$\left\{\sum_i \frac{\Psi_i[\tau+\tau_i;w(t)]}{\Phi_i[\tau+\tau_i;w(t)]}\right\} \times \prod_{j=0,1,2,\ldots} \frac{\Phi_j[\tau+\tau_j;w(t)]}{\Phi_j[\tau_j;w(t)]} \tag{4.75}$$

and the conditional probability of having a specific event

$$\frac{\Psi_i[\tau+\tau_i;w(t)]}{\Phi_i[\tau+\tau_i;w(t)]} \Bigg/ \sum_i \frac{\Psi_i[\tau+\tau_i;w(t)]}{\Phi_i[\tau+\tau_i;w(t)]}\,. \tag{4.76}$$

Notice that all the quantities here are history-dependent, while, if we remove the general dependence on $w(t)$ whilst maintaining the dependence on the elapsed times $\tau_0, \tau_1, \tau_2, \ldots$, we obtain the case of a process with semi-Markov triggers [21]. We now assume that only the left boundary has a non-exponential WTD[6], while the particle triggers have exponential WTDs with rate 1 for free particles in the bulk, and rate β for the particle on the rightmost site. Exponential triggers allow us to write

$$\frac{\Psi_{\mathrm{b}}[\tau+\tau_i;w(t)]}{\Phi_{\mathrm{b}}[\tau+\tau_i;w(t)]} = 1 \tag{4.77}$$

for each free particle in the bulk, while we have

$$\frac{\Psi_L[\tau+\tau_L;w(t)]}{\Phi_L[\tau+\tau_L;w(t)]} = \beta \tag{4.78}$$

for the particle on the rightmost site. Consequently the inter-event time density distribution, conditioned on a time τ_0 having elapsed since the last arrival event, is

$$\begin{aligned}\psi[\tau;w(t)|\tau_0] = &\left(\frac{\Psi_0[\tau+\tau_0;w(t)]}{\Phi_0[\tau_0;w(t)]}(1-\eta_0) + \mathsf{n} + \beta\eta_L\right) \\ &\times \exp\left\{\ln\left[\frac{\Phi_0[\tau+\tau_0;w(t)]}{\Phi_0[\tau_0;w(t)]}\right](1-\eta_0) - (\mathrm{n}+\beta\eta_L)\tau\right\},\end{aligned} \tag{4.79}$$

[6] As we wish to model arrivals with current-dependent rate.

while the probability mass distribution, conditioned on an age τ and elapsed time τ_0 is

$$p_0[\tau; w(t)|\tau_0] = \frac{\Psi_0[\tau+\tau_0; w(t)]}{\Phi_0[\tau_0; w(t)]}(1-\eta_0) \times \left(\frac{\Psi_0[\tau+\tau_0; w(t)]}{\Phi_0[\tau_0; w(t)]}(1-\eta_0) + \mathsf{n} + \beta\eta_L\right)^{-1}, \tag{4.80}$$

$$p_{\mathrm{b}}[\tau; w(t)|\tau_0] = \left(\frac{\Psi_0[\tau+\tau_0; w(t)]}{\Phi_0[\tau_0; w(t)]}(1-\eta_0) + \mathsf{n} + \beta\eta_L\right)^{-1}, \tag{4.81}$$

$$p_L[\tau; w(t)|\tau_0] = \beta\eta_L \left(\frac{\Psi_0[\tau+\tau_0; w(t)]}{\Phi_0[\tau_0; w(t)]}(1-\eta_0) + \mathsf{n} + \beta\eta_L\right)^{-1}, \tag{4.82}$$

where the η_0 and η_L encode the exclusion rules at the boundaries, $i = 1, 2, \ldots, \mathsf{n}$, and n is the number of free particles in the bulk, which depends on the lattice configuration before the jump. The survival probability is

$$\phi[\tau; w(t)|\tau_0] = \exp\left\{\ln\left[\frac{\Phi_0[\tau+\tau_0; w(t)]}{\Phi_0[\tau_0; w(t)]}\right](1-\eta_0) - (\mathrm{n} + \beta\eta_L)\tau\right\}, \tag{4.83}$$

which is the product of the Markovian exponential decays for the bulk and right-boundary particles and a memory-dependent prefactor for the left-boundary arrivals.

We now define the stochastic dynamics of the boundary trigger. Let us impose that the arrival rate α depends linearly on the inwards current $j(t)$, i.e.,

$$\alpha(j) = \alpha_0 + aj, \tag{4.84}$$

which defines a time-dependent rate $\beta_0(t) := \alpha[j(t)]$. A similar functional dependence (but on the instantaneous output current) has been used to model ribosome recycling in protein translation [19, 39]. Generically, such rates describe a simple form of positive feedback (for $a > 0$), whose effect on the stationary state of the TASEP is to shrink the *low-density* phase [23, 39]. The current fluctuations are also altered; the rate function $\hat{e}(j)$ in this phase has already been computed, for our model, by means of the so-called *temporal additivity principle* [23, 25], hence this model provides a testing ground for the cloning method of section 4.2. The particle arrival mechanism starts when the leftmost site is emptied, when we set an age of $\tau = 0$. Denoting by q the current immediately after the last arrival, which occurred at $t - \tau_0$, the value $j(t+\tau)$ at age τ can be expressed as $q(t-\tau_0)/(t+\tau)$, hence the trigger hazard is

$$\beta_0(t+\tau) = \alpha_0 + aq(t-\tau_0)/(t+\tau), \tag{4.85}$$

where τ is the trigger age. Initial values of τ_0 and q are chosen to be 1 and 0, respectively. This allows us to derive the trigger survival probability and the resident time distribution

which are, respectively,

$$\Phi_0[\tau; w(t)] = \exp\left(-\int_0^\tau \beta_0(t+u)\mathrm{d}u\right) \tag{4.86}$$

$$= \left(\frac{t}{t+\tau}\right)^{aq(t-\tau_0)} \mathrm{e}^{-\alpha_0\tau}, \tag{4.87}$$

and $\Psi_0[\tau; w(t)] = \beta_0(t+\tau)\Phi_0[\tau; w(t)]$. Using these in equations (4.79)–(4.82) allows us to generate the trajectories. The survival probability (4.83) for whole process obtained pooling the single-trigger processes has the simple form

$$\Phi_0[\tau; w(t)] = \left(\frac{t}{t+\tau}\right)^{aq(t-\tau_0)(1-\eta_0)} \exp\left\{-\tau\left[\alpha_0(1-\eta_0) - \mathsf{n} - \beta\eta_L\right]\right\}, \tag{4.88}$$

which has been used to generate random inter-event times using the "inverse sampling" method (see, e.g., reference [41]) as follows. With the definitions $Q := aq(t-\tau_0)(1-\eta_0)$ and $R := \alpha_0(1-\eta_0) - \mathsf{n} - \beta\eta_L$, we are concerned with the solution in τ of

$$\left(\frac{t}{t+\tau}\right)^{Q} \mathrm{e}^{-\tau R} = 1-u, \tag{4.89}$$

where u is a drawn from a uniform distribution on $[0,1)$. Equation (4.89) can be rewritten as

$$(t+\tau)R/Q\,\mathrm{e}^{(t+\tau)R/Q} = (1-u)^{-\frac{1}{Q}}\,tR/Q\,\mathrm{e}^{tR/Q} =: C, \tag{4.90}$$

which can be solved numerically with respect to $z := (t+\tau)R/Q$ to get the time to the next jump τ. Interestingly we recognise that $z = W_0(C)$ is the principal branch of the Lambert function [14]. A numerical problem here is that equation (4.90) contains the exponential of t, which for large time causes computational overflow. This has been avoided by taking the logarithm of both sides of equation (4.90) and solving $\ln z + z = \ln C$ with respect to z.

We evaluated the SCGF of the left-boundary current simply by applying a bias e^{-s} to the conditional WTD of the arrival events and implementing the algorithm of section 4.2 for this model. The results are plotted in figure 4.5(a). It is worth noting the existence of a linear branch with slope j^* for large negative values of s. In agreement with the discussion of section 1.6, this branch is mapped to a unique point of the rate function $\hat{e}(j)$ in figure 4.5(b). It would be interesting to probe larger negative values of s but, far from the central regime $s = 0$, the algorithm shows some convergence problems and systematic errors. These indeed are a feature typical of the cloning approaches and are presumably due to the finiteness of the clones population [26]. A simple way proposed

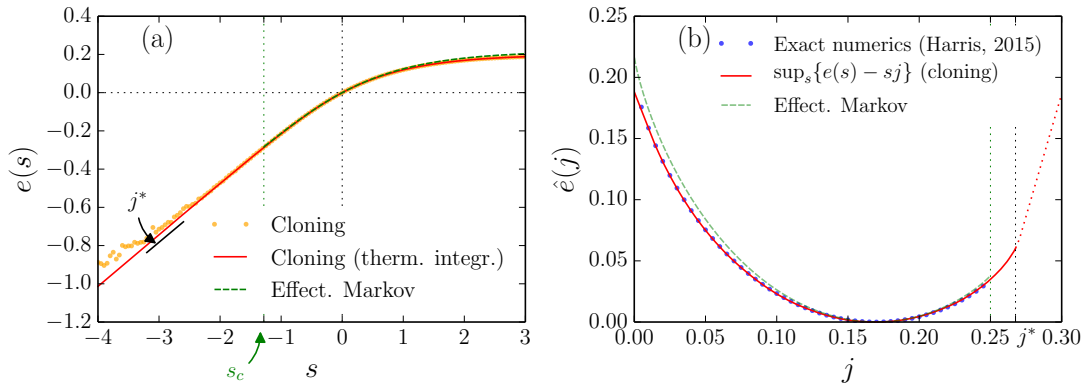

Figure 4.5: (a) Cloning evaluation of the SCGF for the non-Markovian TASEP, with $(\alpha_0, a, \beta, L) = (0.2, 0.1, 1, 10^3)$, using $T = 10^3$. Ensemble size is $N = 5 \cdot 10^3$ ($N = 10^4$) for $s > -2$ ($s < -2$). The markers correspond to the direct evaluation of $e(s)$. Numerical errors are of the order of the symbol size, except for large negative s, where finite-ensemble errors still seem to play a role. The red line is obtained as $\int_0^s (\mathrm{d}e(\sigma)/\mathrm{d}\sigma)\,\mathrm{d}\sigma$, according to the thermodynamic integration of [29]. (b) Comparison between the Legendre–Fenchel transform of the red line in (a) and the rate function of [23]. The dotted line is a numerical artefact due to the finite range of s in (a); the Legendre–Fenchel transform maps the whole linear branch of $e(s)$ to the value at j^* and larger values of j are, in fact, not probed. In both (a) and (b), the dashed lines are the large deviation functionals of the corresponding memory-less TASEP with "effective" arrival rate.

by Lecomte and Tailleur [29] to smooth out the numerical error consists of the following trick. From the definition of the SCGF one gets, by differentiation,

$$\frac{\mathrm{d}}{\mathrm{d}s}e(s) = \frac{1}{Z(s,t)}\int \frac{J[w(t)]}{T}\mathrm{e}^{-sJ[w(t)]}\varrho[w(t)]\mathrm{d}w(t), \tag{4.91}$$

which is the expectation value of the time-averaged current $J[w(t)]/T$ among the population of clones; we can use, of course,

$$e(s) = \int_0^s \left(\frac{\mathrm{d}e(\sigma)}{\mathrm{d}\sigma}\right)\mathrm{d}\sigma, \tag{4.92}$$

thus computing the SCGF from its derivative. Such a procedure is referred to as the *thermodynamic integration* and has been used to make a better estimate of $e(s)$, see the red solid line in figure 4.5(a). A further observation we wish to make here is that the integral in equation (4.92) averages the systematic finite-ensemble error of $e(s)$ from 0 to s; as the error in the neighbourhood of $s = 0$ is negligible, the accuracy of the SCGF at s is enhanced.

If $e(s)$ remains linear with the same slope for $s < -4$, its LF transform will be defined only for $j \leq j^*$. This appears to be related to the dynamical phase transition seen in the Markovian TASEP, where large current fluctuations require correlations on the scale of the system size, and the rate function diverges with L [28]. Indeed the space–time diagrams of the density profile from the cloning simulations, plotted in figure 4.6, seem to suggest that the correlation length increases as s becomes more negative.

In figure 4.5(b), the numerical LF transform of $e(s)$ is validated by the exact numerical minimization calculation of Harris [23], which assumes the temporal additivity principle [25].

Figure 4.5 also shows the large deviation functionals of a standard (Markovian) TASEP obtained by replacing the history dependent arrival rate with a constant "effective" rate $\alpha(\bar{j})$, where $\bar{j}$ is the stationary current and is heuristically obtained as follows. We use the fact that, for the standard TASEP with open boundaries in the low-density phase, the mean current in the long-time limit is given by the arrival rate at the left boundary [28]. Similarly, for the history-dependent TASEP with rate (4.84), intuitively, the current in the long-time limit can be obtained as the solution $\bar{j}$ of the fixed-point equation $j = \alpha(j)$ [23]. The SCGF for the standard TASEP with open boundaries is known analytically [28]; the limit as $L \to \infty$, which represents a valid approximation for $L = 1000$, is

$$e(s) = [1 - \alpha(\bar{j})]\alpha(\bar{j})\frac{1 - \mathrm{e}^{-s}}{1 - \alpha(\bar{j}) + \alpha(\bar{j})\mathrm{e}^{-s}}, \tag{4.93}$$

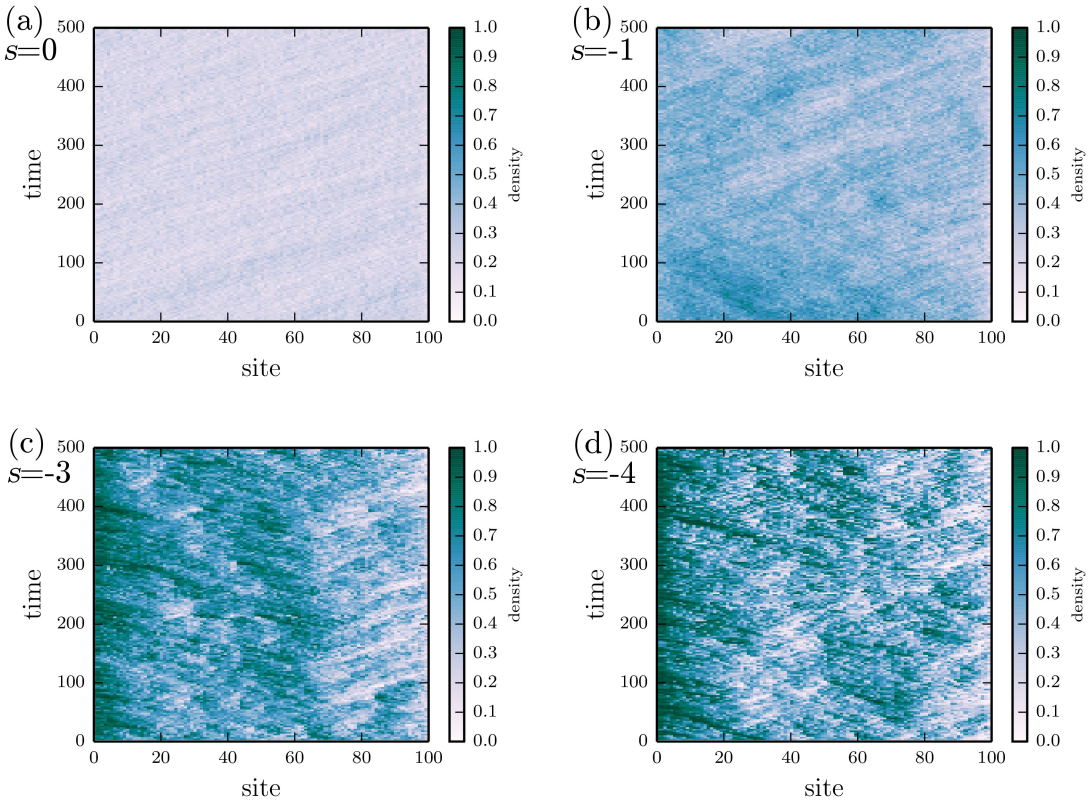

Figure 4.6: Space–time diagrams of the biased history-dependent TASEP of 100 sites. For all cases, the ensemble size is $N = 100$ and the parameters values are $(\alpha_0, a, \beta) = (0.2, 0.1, 1)$. It is possible to appreciate that the density profile is rather uniform in (a) for $s = 0$, while it displays patterns in (b), (c), and (d) for $s < 0$, suggesting an increase of the correlation length with large negative values of s.

for $s > -\ln\left(\frac{1-\alpha(\bar{j})}{\alpha(\bar{j})}\right) =: s_c$, and is plotted in figure 4.5(a) as a dashed line. Its LF transform yields the rate function $\hat{e}(j)$ in the semi-interval $j < 1/4$, also plotted as dashed line in figure 4.5(b). For values of j larger than the critical current $1/4$, $\hat{e}(j)$ is instead proportional to L [28] and can be assumed infinite for a macroscopic chain lattice. We notice that the fluctuations around the mean current for the two models are similar, while it is possible to appreciate that the large deviation functionals differ far from the central regimes. Interestingly, according to the cloning method, the critical current for the history-dependent TASEP j^* is larger than the critical value $1/4$ for the standard "effective" TASEP and it would be interesting to further explore this memory effect.

4.5 Discussion

We have demonstrated that the cloning algorithm for the evaluation of large deviations can be applied consistently for both Markovian *and* non-Markovian dynamics. In fact, the cloning/pruning of trajectories at each temporal step can be performed according to a very simple factor multiplying the WTDs, as in equation (4.2). Our analysis encompasses classes of systems with different memory dependence and exploits the similarities between their different formalisms. The efficacy of this approach is confirmed by numerical results for some of the rare non-Markovian models whose large deviation functions can be obtained exactly.

For general non-Markovian cases, the implementation of our procedure is not much harder than the exact simulation of the original trajectories. In Markov processes, the procedure is equivalent to those of Giardinà et al. [18], Lecomte and Tailleur [29], where biased dynamics involving alternative rates or transition probabilities have been defined. We expect that, to minimize finite-ensemble effects, an optimal choice of modified WTDs and cloning factors exists for both non-Markovian and Markovian systems, along the lines of the feedback control of Nemoto et al. [35]. Further developments can thus be anticipated. We also mention that the discrete-time case of Giardinà et al. [18] is interesting as the jumps and the cloning steps occur simultaneously for each ensemble element. This feature can be used to prevent a single clone replacing a macroscopic fraction of the ensemble, thus reducing finite size effects.

Large deviation functionals are often hard to obtain analytically, and such a difficulty is exacerbated in non-Markovian systems, which better describe real-world situations. We think that the results of this work open up a promising avenue for numerical studies.

To explore real-world applications of the cloning method, in the next chapter we

consider some aspects of teletraffic engineering, where time-extensive observables and their large deviations are of special interest.

Bibliography

[1] D. F. Anderson. A modified next reaction method for simulating chemical systems with time dependent propensities and delays. *J. Chem. Phys.*, 127(21):214107, 2007.

[2] D. Andrieux and P. Gaspard. The fluctuation theorem for currents in semi-Markov processes. *J. Stat. Mech.*, 2008(11):P11007, 2008.

[3] E. Barkai, R. S. Eisenberg, and Z. Schuss. Bidirectional shot noise in a singly occupied channel. *Phys. Rev. E*, 54(2):1161–1175, 1996.

[4] D. Ben-Avraham and S. Havlin. *Diffusion and Reactions in Fractals and Disordered Systems*. Cambridge University Press, Cambridge UK, 2000.

[5] J. T. Berryman and T. Schilling. Sampling rare events in nonequilibrium and non-stationary systems. *J. Chem. Phys.*, 133(24):244101, 2010.

[6] M. Boguñá, L. F. Lafuerza, R. Toral, and M. Á. Serrano. Simulating non-Markovian stochastic processes. *Phys. Rev. E*, 90(4):1–9, 2014.

[7] D. Bratsun, D. Volfson, L. S. Tsimring, and J. Hasty. Delay-induced stochastic oscillations in gene regulation. *Proc. Natl. Acad. Sci. USA*, 102(41):14593–8, 2005.

[8] T. Brett and T. Galla. Stochastic processes with distributed delays: Chemical langevin equation and linear-noise approximation. *Phys. Rev. Lett.*, 110(25):250601, 2013.

[9] X. Cai. Exact stochastic simulation of coupled chemical reactions with delays. *J. Chem. Phys.*, 126(12):124108, 2007.

[10] M. Cavallaro and R. J. Harris. A framework for the direct evaluation of large deviations in non-Markovian processes. 2016. To appear as a Letter in J. Phys. A: Math. Theor.

[11] M. Cavallaro, R. J. Mondragón, and R. J. Harris. Temporally correlated zero-range process with open boundaries: Steady state and fluctuations. *Phys. Rev. E*, 92(2): 022137, 2015.

[12] D. Chiarugi, M. Falaschi, D. Hermith, C. Olarte, and L. Torella. Modelling non-Markovian dynamics in biochemical reactions. *BMC Syst, Biol.*, 9 Suppl 3(Suppl 3):S8, 2015.

[13] R. J. Concannon and R. A. Blythe. Spatiotemporally complete condensation in a non-Poissonian exclusion process. *Phys. Rev. Lett.*, 112(5):050603, 2014.

[14] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth. On the LambertW function. *Adv. Comput. Math.*, 5(1):329–359, 1996.

[15] P. S. Efraimidis and P. G. Spirakis. Weighted random sampling with a reservoir. *Inf. Process. Lett.*, 97(5):181–185, 2006.

[16] M. Esposito and K. Lindenberg. Continuous-time random walk for open systems: Fluctuation theorems and counting statistics. *Phys. Rev. E*, 77(5):051119, 2008.

[17] J. P. Garrahan, R. L. Jack, V. Lecomte, E. Pitard, K. van Duijvendijk, and F. van Wijland. First-order dynamical phase transition in models of glasses: an approach based on ensembles of histories. *J. Phys. A: Math. Gen.*, 42(7):075007, 2009.

[18] C. Giardinà, J. Kurchan, and L. Peliti. Direct evaluation of large-deviation functions. *Phys. Rev. Lett.*, 96(12):120603, 2006.

[19] M. A. Gilchrist and A. Wagner. A model of protein translation including codon bias, nonsense errors, and ribosome recycling. *J. Theor. Biol.*, 239(4):417–34, 2006.

[20] K.-I. Goh and A.-L. Barabási. Burstiness and memory in complex systems. *EPL*, 81(4):48002, 2008.

[21] M. Gorissen and C. Vanderzande. Ribosome Dwell Times and the Protein Copy Number Distribution. *J. Stat. Phys.*, 148(4):627–635, 2012.

[22] M. Gorissen, J. Hooyberghs, and C. Vanderzande. Density-matrix renormalization-group study of current and activity fluctuations near nonequilibrium phase transitions. *Phys. Rev. E*, 79(2 Pt 1):020101, 2009.

[23] R. J. Harris. Fluctuations in interacting particle systems with memory. *J. Stat. Mech.*, 2015(7):P07021, 2015.

[24] R. J. Harris. Random walkers with extreme value memory: modelling the peak-end rule. *New J. Phys.*, 17(5):053049, 2015.

[25] R. J. Harris and H. Touchette. Current fluctuations in stochastic systems with long-range memory. *J. Phys. A: Math. Gen.*, 42(34):342001, 2009.

[26] P. I. Hurtado and P. L. Garrido. Current fluctuations and statistics during a large deviation event in an exactly solvable transport model. *J. Stat. Mech.*, 2009(02): P02032, 2009.

[27] D. Khoromskaia, R. J. Harris, and S. Grosskinsky. Dynamics of non-Markovian exclusion processes. *J. Stat. Mech.*, 2014(12):P12013, 2014.

[28] A. Lazarescu. The physicist's companion to current fluctuations: one-dimensional bulk-driven lattice gases. *J. Phys. A: Math. Gen.*, 48(50):503001, 2015.

[29] V. Lecomte and J. Tailleur. A numerical approach to large deviations in continuous time. *J. Stat. Mech.*, 2007(03):P03004–P03004, 2007.

[30] C. Maes, K. Netočný, and B. Wynants. Dynamical fluctuations for semi-Markov processes. *J. Phys. A: Math. Gen.*, 42(36):365002, 2009.

[31] Rocha L. E. C. Masuda N. A Gillespie algorithm for non-Markovian stochastic processes: Laplace transform approach. 2016.

[32] M. Merolle, J. P. Garrahan, and D. Chandler. Space-time thermodynamics of the glass transition. *Proc. Nat. Ac. Sci. USA*, 102(31):10837–10840, 2005.

[33] J. D. Murray. *Mathematical Biology: I. An Introduction.* Springer-Verlag, New York NY, 2007.

[34] T. Nemoto and S. Sasa. Computation of Large Deviation Statistics via Iterative Measurement-and-Feedback Procedure. *Phys. Rev. Lett.*, 112(9):090602, 2014.

[35] T. Nemoto, F. Bouchet, R. L. Jack, and V. Lecomte. Population-dynamics method with a multicanonical feedback control. *Phys. Rev. E*, 93(6):062123, 2016.

[36] F. Paraan and J. Esguerra. Exact moments in a continuous time random walk with complete memory of its history. *Phys. Rev. E*, 74(3):032101, 2006.

[37] J. Realpe-Gomez, T. Galla, and A. J. McKane. Demographic noise and piecewise deterministic Markov processes. *Phys. Rev. E*, 86(1):011137, 2012.

[38] G. M. Schütz and S. Trimper. Elephants can always remember: Exact long-range memory effects in a non-Markovian random walk. *Phys. Rev. E*, 70(4):045101, 2004.

[39] A. K. Sharma and D. Chowdhury. Stochastic theory of protein synthesis and polysome: Ribosome profile on a single mRNA transcript. *J. Theor. Biol.*, 289: 36 – 46, 2011.

[40] R. D. Smith. The dynamics of internet traffic: self-similarity, self-organization, and complex phenomena. *Adv. Complex Syst.*, 14(06):905–949, 2011.

[41] W. J. Stewart. *Probability, Markov Chains, Queues, and Simulation: The Mathematical Basis of Performance Modeling.* Princeton University Press, Princenton NJ, 2009.

[42] S. Trimper, K. Zabrocki, and M. Schulz. Memory-controlled diffusion. *Phys. Rev. E*, 70(5):056133, 2004.

[43] P. Van Mieghem and R. van de Bovenkamp. Non-Markovian infection spread dramatically alters the susceptible-infected-susceptible epidemic threshold in networks. *Phys. Rev. Lett.*, 110(10):108701, 2013.

[44] J. Voit. *The Statistical Mechanics of Financial Markets.* Texts and Monographs in Physics. Springer-Verlag, Berlin Heidelberg Germany, second edition, 2005.

5 Applications to telecommunications engineering

Contents

Natural systems made of many coupled components, ranging from ideal gases to living organisms and their communities, have long been of interest to scientists. By contrast, recently, some of the most studied complex systems are man-made, for instance telecommunication networks and financial markets. The ways used to approach these technological systems is similar to those used in natural sciences. Indeed, at a sensible level of detail, the system's properties appear as random variables, and the scientist effort is directed towards the quantification of such randomness, as well as of its effects. Specifically, in telecommunication engineering, we are interested in relating the elementary ("microscopic") description of the telecommunication networks in terms of packets and servers, to a perceivable ("macroscopic") quantity, such as the service available to the final user.

Here, indeed, we aim at demonstrating that it is possible to apply the machinery developed in the previous chapters to a real-world problem in telecommunication engineering, that is, "how can we decide whether a server can accept or refuse some work,

without dissatisfying existing customers?". We only deal with simple examples, while there is potentially much more which could be done in terms of actual applications.

This small chapter is organised as follows. In section 5.1 the concept of *workload* is introduced. In section 5.2 we state three theorems that allow us to bound probabilities. By means of these, we introduce and motivate the notion of *effective bandwidth* in section 5.3, which is closely related to that of the SCGF. Finally, in section 5.4, we consider toy models of packet traffic and some analogies with the formalism developed in the previous chapter, showing how to use the cloning method of chapter 4 to compute their effective bandwidths. We conclude the chapter in section 5.5.

5.1 Introduction

In a queuing network, we have a collection of servers that exchange packets or customers. In chapter 2 we drew a parallel between the ZRP and a queueing network, but in order to study more general non-Markovian queuing systems, we find convenient to set up a slightly different notation than the previous chapters, based on references [14, 15, 18]. When a collection of packets leaves a server to reach another one, we say that a communication channel has been established. In such a channel, we refer to the random amount of work brought by customers arriving during the interval $[t_0, t)$ as $X(t_0, t)$. This is the integral over $[t_0, t)$ of a renewal process $x(t)$, viz., a series of random events which describe particles (or customers) feeding a server. Similarly we consider the amount of work $Y(t_0, t)$ that the server can do during the same amount of time, which is the integral over $[t_0, t)$ of another renewal process $y(t)$. It is of central importance here that we assume that such quantities are time-extensive. They can thus play the role of the current $J[w(t)]$, as defined in the previous chapters of this thesis, and their rare statistics can be studied using the cloning method of chapter 4.

We now define the random variable $W(t_0, t) = X(t_0, t) - Y(t_0, t)$, which is referred to as the *workload process*. As shown in figure 5.1, $W(t_0, t)$ has an increment at the instants $(t_1, t_2, \ldots, t_n)$, with $t_0 \leq t_1 \leq \ldots \leq t_n < t$, where an arrival or service event occurs. It is important to notice the service cannot be stored. After each increment, it is possible that there still is a certain amount of work waiting to be done; this is referred to as the queue length and is denoted by $Q(t)$. The dynamics of the server are as follows. The work to be done at time t_n is the sum of the length of the queue at the previous event instant t_{n-1} and the workload increment $W(t_0, t_n) - W(t_0, t_{n-1})$ done during the same interval; however, when $W(t_0, t_n) - W(t_0, t_{n-1}) + Q(t_{n-1}) < 0$, the work surplus is wasted, yielding a zero queue length (instead of a negative one). This can all be expressed compactly in

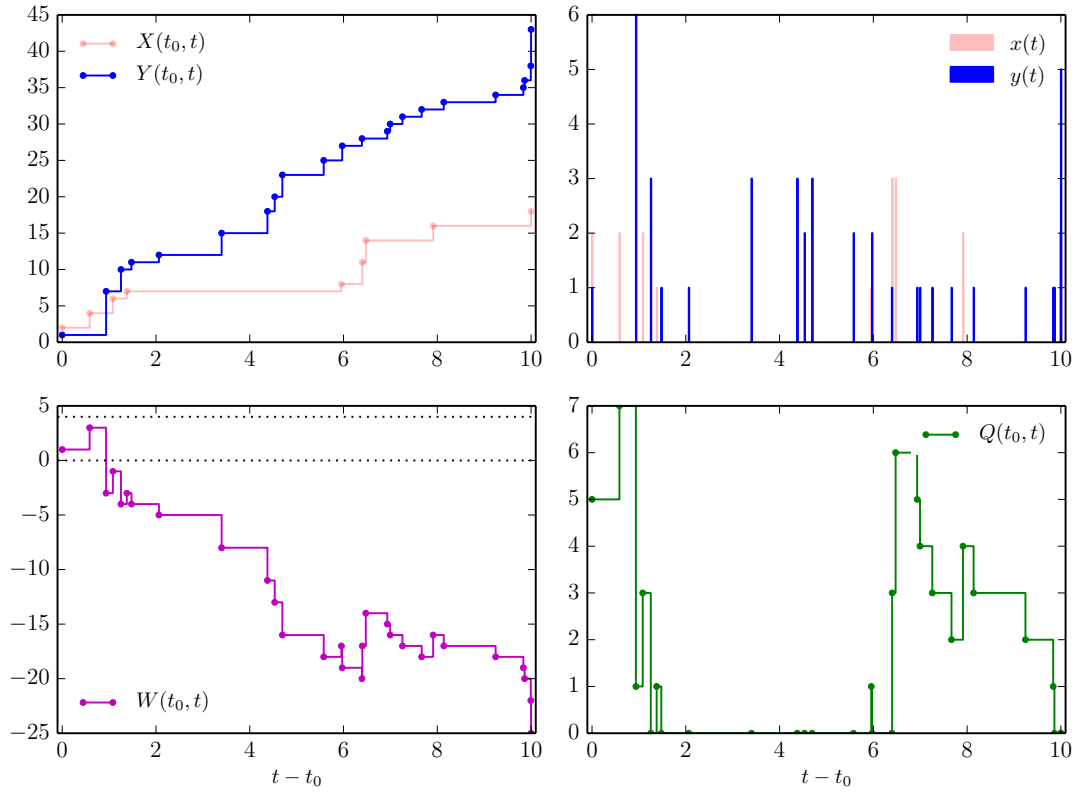

Figure 5.1: Workload process. (a) The light red line is the total amount of work (e.g., customers to be served) requested during the interval $[t_0, t)$, while the blue line is the service that can be provided during the same amount of time. (b) The red spikes are the arrival events, while the blue spikes are the attempted service events. (c) Workload process $W(t_0, t) = X(t_0, t) - Y(t_0, t)$. (d) Time evolution of the occupation of a queue subjected to the arrivals and services of (a) and (b). The initial occupation number is 4.

the recursive relation

$$Q(t_n) = \max\{0, W(t_0, t_n) - W(t_0, t_{n-1}) + Q(t_{n-1})\}. \tag{5.1}$$

In a discrete time setting, the events can be thought of being equally spaced and equation (5.1) is referred to as Lindley's formula. The definition of $Q(t)$ given the workload process when both time and packet sizes are continuous is much more subtle and can be found, e.g., in reference [11]. A simplification consists of assuming constant deterministic service (i.e., a queue of type G/D/1, according to the Kendall notation). In the next sections, we will deal with such a type of queue, unless explicitly specified.

5.2 Bounds and limit theorems

A very loose bound theorem on a generic random variable X that takes only nonnegative values with density function $f(u)$ can be derived from the knowledge of its expectation value, i.e.,

$$\int_0^\infty uf(u)\,\mathrm{d}u \geq x \int_x^\infty f(u)\,\mathrm{d}u, \tag{5.2}$$

for $x \geq 0$. This can be rewritten more conveniently as

$$\mathrm{Prob}\{X \geq x\} \leq \frac{\langle X \rangle}{x}. \tag{5.3}$$

A more general version of the bound (5.2) valid for nonnegative and nondecreasing functions h of X is called *Markov inequality* and reads,

$$\int_{-\infty}^\infty h(u)f(u)\,\mathrm{d}u \geq h(x) \int_x^\infty f(u)\,\mathrm{d}u. \tag{5.4}$$

The Markov inequality is not immediately useful for many real-world situations (as the bound can be very loose) but allows us to derive more interesting bounds as follows. We assume that the variance $\sigma_X^2 = \langle (X - \langle X \rangle)^2 \rangle$ of X is finite, and define a new random variable

$$Y = (X - \langle X \rangle)^2. \tag{5.5}$$

Then, from relation (5.3), we get $\mathrm{Prob}\{Y \geq x^2\} \leq \frac{\langle Y \rangle}{x^2}$, which can also be written as

$$\mathrm{Prob}\{|X - \langle X \rangle| \geq x\} \leq \frac{\sigma_X^2}{x^2}, \tag{5.6}$$

and is referred to as the *Chebychev inequality* for X. Notice that this is more general

that equation (5.3) as it holds for random variables that can take negative values.

We now consider a function $h(x) = \mathrm{e}^{sx}$, for $s > 0$ in equation (5.4) and obtain

$$\mathrm{Prob}\{X \geq x\} \leq \mathrm{e}^{-sx}\langle \mathrm{e}^{sX}\rangle. \tag{5.7}$$

This inequality is referred to as the *Chernoff bound*. Further details can be found, e.g., in the book of W. J. Stewart [26]. In the next section we see that it provides an useful link between the theory developed in chapters 1 and 4 and the assessment of the performance of a communication channel.

5.3 Effective bandwidth

We refer to the *effective bandwidth* $\mathsf{e}(s,t)$ of a process $X(t_0,t)$ of duration T that starts at t_0, ends at $t = T + t_0$, and describes a time-extensive observable, as the functional

$$\mathsf{e}(s,t) = \frac{\ln\langle \mathrm{e}^{sX(t_0,t)}\rangle}{s(t-t_0)}. \tag{5.8}$$

This notion provides a way to evaluate the performance of a communication channel subjected to an arbitrary arrival process, as explained, e.g., in references [14, 15] and outlined below. When the service is continuously and deterministically provided with a constant rate r, the service capacity in the time interval $t - t_0$ is $r(t - t_0)$, and the total workload process for this amount of time is $W(t_0,t) = X(t_0,t) - r(t-t_0)$. In telecommunications, the quantity r is typically a data transfer rate and it is referred to as the *bandwidth*. We also assume that a number n of sources contributes to $X(t_0,t)$, i.e.,

$$X(t_0,t) = \sum_{i=0}^{n} X_i(t_0,t) \tag{5.9}$$

According to the Chernoff bound (5.7), the probability that the service request overflows the capacity satisfies

$$\ln \mathrm{Prob}\{X(t_0,t) > r(t-t_0)\} \leq \ln\langle \mathrm{e}^{s(X(t_0,t)-r(t-t_0))}\rangle \tag{5.10}$$

$$= \ln\langle \mathrm{e}^{s(X(t_0,t))}\rangle - s\,r(t-t_0). \tag{5.11}$$

for all $s > 0$. In this context, we say that a certain *quality of service*—for which the acronym QoS is commonly used—for a given γ, is guaranteed if the following condition

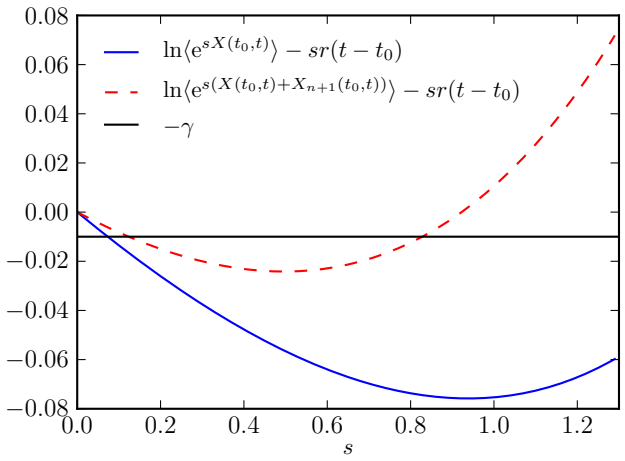

Figure 5.2: QoS control. $X(t_0,t)$ and $X_{n+1}(t_0,t)$ are Poisson processes of duration $(t-t_0)$ with rates 0.09 and 0.05, respectively. The QoS is calibrated on $\gamma = -0.01$ and the deterministic service rate is $r = 0.23$. As there are values of s such that the moment generating function of $X(t_0,t) + X_{n+1}(t_0,t)$ is smaller than $sr(t-t_0) - \gamma$, the condition (5.12) for $X(t_0,t) + X_{n+1}(t_0,t)$ is satisfied and the new arrival X_{n+1} can be accepted.

is satisfied:

$$\text{Prob}\{X(t_0,t) > r\,(t-t_0)\} \leq \mathrm{e}^{-\gamma}. \tag{5.12}$$

The same inequality is also referred to as *service requirement* or *admission criteria*. We prefer to work with another condition instead,

$$\inf_{s>0}\{\ln\langle \mathrm{e}^{s(X(t_0,t))}\rangle - s\,r\,(t-t_0)\} \leq -\gamma, \tag{5.13}$$

which is sufficient for equation (5.12) to hold. This means that if $\ln\langle \mathrm{e}^{s(X(t_0,t))}\rangle - sr(t-t_0)$ is less than $-\gamma$ *for some* $s > 0$, then the QoS is guaranteed. We are now in the position to decide whether the server can accept another service request $X_{n+1}(t_0,t)$ (which is independent of $X(t_0,t)$), without violating the condition (5.12). The criterion is that the new request $X_{n+1}(t_0,t)$ is accepted if there is at least a value s such that

$$\ln\langle \mathrm{e}^{s(X(t_0,t))}\rangle + \ln\langle \mathrm{e}^{s(X_{n+1}(t_0,t))}\rangle - sr(t-t_0) \leq -\gamma. \tag{5.14}$$

A simple example where a new Poisson arrival process $X_{n+1}(t_0,t)$ can be accepted is shown in figure 5.2. Dividing by $s(t-t_0)$, we get

$$\frac{\ln\langle \mathrm{e}^{s(X(t_0,t))}\rangle}{s(t-t_0)} + \frac{\ln\langle \mathrm{e}^{s(X_{n+1}(t_0,t))}\rangle}{s(t-t_0)} \leq r - \frac{\gamma}{s(t-t_0)}. \tag{5.15}$$

This explicitly shows how the effective bandwidths compare with the true bandwidth r. In conclusion, we can evaluate whether a new connection allows us to maintain a promised QoS by computing the effective bandwidths of the sources $X(t, t_0)$ and $X_{n+1}(t, t_0)$. Clearly, the effective bandwidth can be calculated for the outwards traffic departing from a server, as well as for the available work $Y(t_0, t)$. In the next sections we will compute the effective bandwidths of selected processes and we will use the cloning method detailed in chapter 4.

5.4 Examples

5.4.1 A periodic source

As a first example, we consider a simple source that produces a given (fixed) real positive number b of packets at random times $(n + U_n)d$, where the U_n, $n = 0, 1, 2, \ldots$, are uniformly distributed on the interval $[0, 1)$, and $0 < d < 1$. This model can be thought of a source where b particles are produced at each cycle (the period of each cycle being subjected to noise) and has been used to describe the workload produced by a constant rate information source [23]. The analytic expression for its effective bandwidths appears in reference [14] and is plotted in figure 5.3; it reads:

$$\mathsf{e}(s,t) = \frac{b}{t}\left\lfloor \frac{t}{d} \right\rfloor + \frac{1}{st}\ln\left[1 + \left(\frac{t}{d} - \left\lfloor \frac{t}{d} \right\rfloor\right)(\mathrm{e}^{sb} - 1)\right]. \tag{5.16}$$

We derive equation (5.16) as follows. The arrival process $X(t_0, t)$ can be written as

$$X(t_0, t) = b\left(\left\lfloor \frac{t}{d} \right\rfloor + B(p)\right), \tag{5.17}$$

where $B(p)$ is a Bernoulli random variable with $p = \frac{t}{d} - \left\lfloor \frac{t}{d} \right\rfloor$, i.e., $B(p)$ is zero with probability p and one with probability $q = 1 - p$. The moment generating function of $X(t)$ is simply obtained as the product of the generating functions of a constant and a Bernoulli random variable, viz.,

$$\langle \mathrm{e}^{sX(t_0,t)} \rangle = \exp\left(sb\left\lfloor \frac{t}{d} \right\rfloor\right)(q + p\mathrm{e}^{sb}), \tag{5.18}$$

which, after taking the logarithm and dividing by st, leads to equation (5.16). The same functional can be evaluated directly realising an ensemble of N trajectories according to $X(t_0, t)$, but cloning/pruning some of the trajectories by a factor $(N + \mathrm{e}^s - 1)/N$ at each arrival, as detailed in chapter 4, section 4.2.2. The result obtained with this method is

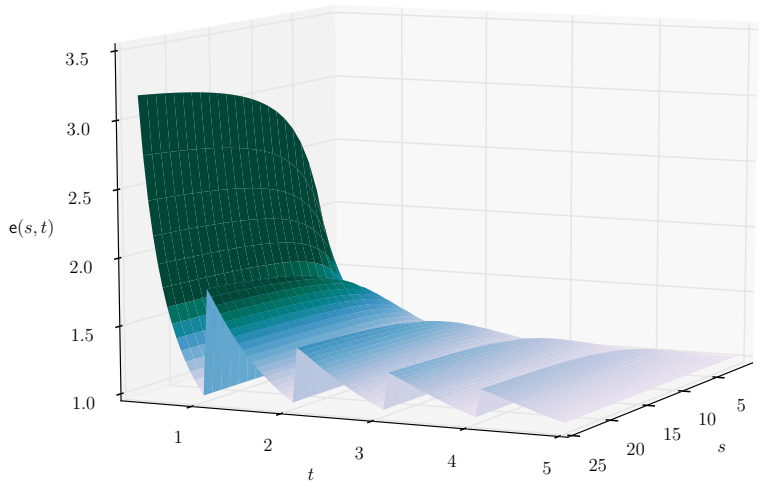

Figure 5.3: Effective bandwidth of the periodic source (5.17) with parameters $b = d = 1$. The source produces a single packet at each step.

compared to the analytical curve in figure 5.4. Not surprisingly, the two approaches are consistent.

5.4.2 Fluid workload

Generically, a fluid queue is a stochastic model used to describe the flow out of a source (or server) subjected to random periods of filling and emptying. It was first introduced by P. A. P. Moran to describe the level of a dam, based on a discrete-time stochastic process [20]. Since then, many continuous-time variants have been used in wildfire modelling [25], in ruin theory [2], and, above all, in high-speed data-networks simulations, see, e.g., reference [1].

A source or workload is called a Markov fluid if its time derivative is function of a continuous-time Markov process. We consider a server that can be in many states, and during the stay on each state it releases a fluid with a certain deterministic rate. In such a model, deterministic and stochastic dynamics coexists, and the traffic generated is represented as a piece-wise continuous flow, which contrasts with the particle/packet models. The flow intensity varies according to the state of an underlying continuous-time Markov process. Each state i of this Markov process corresponds to a different flow b_i.

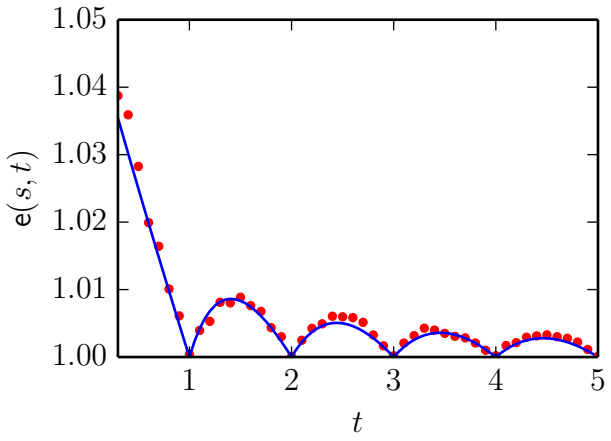

Figure 5.4: Section of the effective bandwidth (5.16) with parameters $b = d = 1$ and $s = 1$. The solid line corresponds to the analytical prediction (5.16). The points correspond to the numerical results using the cloning algorithm of chapter 4.

This generates a workload process which we now call $B(t_0, t)$, to point out the analogy with the type B observables found in chapters 1 and 4. As an example, combining a number n of identical on-off sources, the configuration state of the underlying Markov chain has dimension $|\mathcal{S}| = n$, each one corresponding to the number of active source. The flow is at its peak when all the sources are active. The set of all the possible values of total flow are organised into the matrix $\mathbf{B} = \mathrm{diag}(b_1, b_2, \ldots, b_{|\mathcal{S}|})$. Rather than formally developing the whole theory of fluid queues, focus on the analogies with the formalism seen so far. The total workload process can be written as *continuous* functional of the trajectory $(t_0, x_0,\ t_1, x_1,\ t_2, x_2\ \ldots\ t_n, x_n,\ t)$:

$$B(t_0, t) = \mathrm{B}[w(t)] = \sum_{i=0}^{n-1} (t_{i+1} - t_i) b_{x_i} + (t - t_n) b_{x_n}. \tag{5.19}$$

Similarly to the type B observables, each term of this only depends on the time-increment $(t_{i+1} - t_i)$ and the configuration x_i, excluding the last term. In analogy with what we seen in the previous chapters, where we worked on a joint configuration-current space, we now define a joint configuration-flow space and refer to the probability of having a configuration x_i, at time t, with a value B for the flow as $P_{x_i}(\mathsf{B}, t)$. This probability satisfies the Master equation

$$\frac{\mathrm{d}}{\mathrm{d}t} P_{x_i}(\mathsf{B}, t) = \sum_{x_j} [\mathbf{G}]_{x_i, x_j} P_{x_j}(\mathsf{B}, t) + b_{x_i} \frac{\mathrm{d}}{\mathrm{d}\mathsf{B}} P_{x_i}(\mathsf{B}, t), \tag{5.20}$$

where $\mathbf{G}$ is the stochastic generator of the underlying Markov process. Now, this sys-

tem can be diagonalised with respect to the subspace of Bs, by means of a Laplace transformation that yields the biased Master equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\tilde{P}_{x_i}(t) = \sum_{x_j}[\mathbf{G}]_{x_i,x_j}\tilde{P}_{x_j}(t) + sb_{x_i}\tilde{P}_{x_i}(t), \tag{5.21}$$

which, along the lines of the algorithm of [17], as described in section 1.7.2, corresponds to the dynamics of a system that changes configuration according to $\mathbf{G}$, whose weight evolves exponentially with rate sb_{x_i} during the stay in state x_i. Also, after a visit to x_i of duration τ, the observable $B(t_0,t)$ increased by $b_{x_i}\tau$.

We now present a derivation of the effective bandwidth of a Markov fluid that follows closely the one of reference [16]. We follow a standard argument based on the Master equation (1.14), while reference [16] is based on the backward equation (1.16), which is more popular in the Queuing Theory community. The aim is to find $\mathsf{e}(s,t) = \ln\langle \mathrm{e}^{sB(t_0,t)}\rangle/(s(t-t_0)) = \ln\langle 1|\mathrm{e}^{sB(t_0,t)}|\tilde{P}(t_0)\rangle/(s(t-t_0))$. We focus first on

$$|\tilde{P}(t)\rangle = \mathrm{e}^{sB(t-\mathrm{d}t,t)}|\tilde{P}(t-\mathrm{d}t)\rangle. \tag{5.22}$$

Component-wise, this is equivalent to

$$\tilde{P}_{x_i}(t) = \sum_{x_j}\left[\mathrm{e}^{\mathbf{G}\mathrm{d}t}\right]_{x_i,x_j}\left[\mathrm{e}^{s\mathbf{B}\mathrm{d}t}\right]_{x_j,x_j}\tilde{P}_{x_j}(t-\mathrm{d}t). \tag{5.23}$$

Since $[\exp(\mathbf{G}\mathrm{d}t)]_{x_i,x_j} = [\mathbb{1}]_{x_i,x_j} + [\mathbf{G}]_{x_i,x_j}\,\mathrm{d}t + o(\mathrm{d}t)$ and $[\exp(s\mathbf{B}\mathrm{d}t)]_{x_i,x_i} = 1 + sb_{x_i}\mathrm{d}t + o(\mathrm{d}t)$, we get

$$\frac{\tilde{P}_{x_i}(t) - \tilde{P}_{x_i}(t-\mathrm{d}t)}{\mathrm{d}t} = \tilde{P}_{x_i}(t-\mathrm{d}t)([\mathbf{G}]_{x_i,x_i}+sb_{x_i}) + \sum_{x_j\neq x_i}\tilde{P}_{x_i}(t-\mathrm{d}t)[\mathbf{G}]_{x_j,x_i} + O(\mathrm{d}t) \tag{5.24}$$

In the limit as $\mathrm{d}t\to 0$, we directly get equation (5.21). In matrix form this equation is

$$\frac{\mathrm{d}}{\mathrm{d}t}|\tilde{P}(t)\rangle = (\mathbf{G}+s\mathbf{B})|\tilde{P}(t)\rangle \tag{5.25}$$

and has formal solution

$$|\tilde{P}(t)\rangle = \exp\left[(\mathbf{G}+s\mathbf{B})(t-t_0)\right]|\tilde{P}(t_0), \tag{5.26}$$

from which the following form for the effective bandwidth is obtained

$$\mathsf{e}(s,t) = \frac{1}{st}\ln\langle 1|\exp\left[(\mathbf{G}+s\mathbf{B})(t-t_0)\right]|\tilde{P}(t_0)\rangle. \tag{5.27}$$

As a simple example, we focus now on a simple source modulated by a *telegraph process* i.e., a two-state (0 and 1) continuous-time Markov process with generator

$$\mathbf{G} = \begin{pmatrix} -\alpha & \beta \\ \alpha & -\beta \end{pmatrix}. \tag{5.28}$$

When the configuration is 0, workload is produced deterministically at a rate $b_0 = h$, while no workload is produced during a stay in state 1,

$$b_i = \begin{cases} h & \text{if } i = 0 \\ 0 & \text{if } i = 1 \end{cases}, \tag{5.29}$$

as the example of reference [14]. The biased Master equation explicitly reads

$$\frac{\mathrm{d}}{\mathrm{d}t} \begin{pmatrix} \tilde{P}_0(t) \\ \tilde{P}_1(t) \end{pmatrix} = \begin{pmatrix} -\alpha + hs & \beta \\ \alpha & -\beta \end{pmatrix} \begin{pmatrix} \tilde{P}_0(t) \\ \tilde{P}_1(t) \end{pmatrix}. \tag{5.30}$$

Such a model would describe a source that is either in a idle state, i.e., not transmitting any packet, or in a active state and transmitting at its peak rate. Assuming that the observation starts at the stationary distribution

$$|P^*\rangle = \begin{pmatrix} \frac{\beta}{\beta+\alpha} \\ \frac{\alpha}{\beta+\alpha} \end{pmatrix}, \tag{5.31}$$

we can compute the effective bandwidth

$$\mathsf{e}(s,t) = \frac{1}{st} \ln\langle 1| \exp\left[\begin{pmatrix} -\alpha + hs & \beta \\ \alpha & -\beta \end{pmatrix} t \right] |P^*\rangle. \tag{5.32}$$

We mention that a related model is the so called *Markov-modulated Poisson process* (MMPP) [10], which is a Poisson process with random intensity, defined by an underlying Markov process. This can be still represented by the Master equation (5.25), but has a different interpretation, as each state x_i produces a Poisson process with intensity b_{x_i}. Both fluid sources and MMPP are characterised by the generator $\mathbf{G}$ and the rate diagonal matrix $\mathbf{B}$. Furthermore, one can think of replacing the Poisson process in a MMPP with a generalised birth-death process, thus obtaining a similar but more general process, which has been used to model populations in randomly switching environments (see, e.g., Hufton et al. [13] and references therein). Within such class of models, the lowest-order approximation of the modulated birth-death process leads to the *piecewise-deterministic Markov processes* (PDMPs) [6, 7], which have been recently shown to be appropriate also for the natural sciences (where the underlying Markov process represents

the extrinsic noise) [13, 19, 21, 22, 28, 29]. Such PDMPs are more general than the Markov fluid (as defined earlier in this section), as the transition rates of its underlying Markov process can depend on the state of the population and the deterministic part can be non-linear (while the fluid workload is linear). It is finally worth mentioning that a non-equilibrium statistical mechanics of PDMPs is presented in references [5, 9].

5.4.3 Statistics of packet loss

We present here another simple example of s-modified generators $\tilde{\mathbf{G}}$ with biased rates in the diagonal entries $[\tilde{\mathbf{G}}]_{x_i,x_i}$. Let us consider a one-dimensional random walk on a linear chain of length N. When the walker is in position N, a new arrival (with rate λ) causes the total-current counter to tick, but leaving the occupation number unchanged. Such a system has a lucid interpretation in queuing theory and is referred to as an M/M/1/N queue in the Kendall notation. Customers arrive according to a Poisson process at rate λ and are processed by a single server at rate μ. Contrarily to an M/M/1 queue, here there is space in the server only for N customers. When the server is fully occupied, there is no interruption of the arrival process; however the new customers do not alter the queue, simply disappearing instead. In communication systems, such customers are said to be "lost". Formally, the occupation number of the queue follows a birth-death process, where the new arrivals can be neglected when $n \geq \mathsf{N}$, i.e., $\lambda_n = 0$ for $n \geq \mathsf{N}$, $\lambda_0 = 0$, and $\lambda_n = \lambda$ elsewhere. Hence, the stationary state is

$$P_n^* = \frac{1-\lambda/\mu}{1-(\lambda/\mu)^{\mathsf{N}+1}}(\lambda/\mu)^n, \quad n \leq \mathsf{N}, \tag{5.33}$$

if $\lambda \neq \mu$, or $P_n^* = 1/(\mathsf{N}+1)$, if $\lambda = \mu$. This can be obtained easiliy and is in agreement with the detailed-balance solution (1.101).

We are interested in the statistics of particle loss, i.e., we want to count the number of customers that arrive when the occupation number of the queue is N. For this case, we do not consider the effective bandwidth, but rather compute the standard SCGF. The mean packet loss is simply given by the arrival rate λ times the probability P_{N}^* that the queue is full. Such arrivals correspond to jumps that leave that state as it is, but

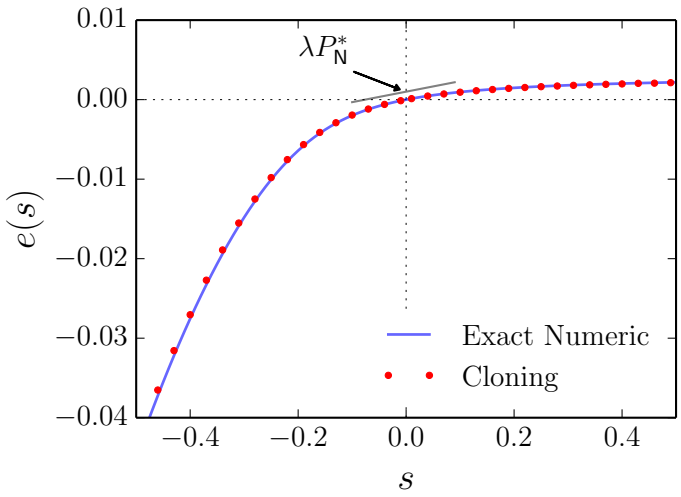

Figure 5.5: SCGF of packet loss in an M/M/N/1 queue with $(\lambda, \mu, \mathsf{N}) = (0.19, 0.20, 10)$. The solid line is the lowest eigenvalue of the matrix (5.34). Esemble size and simulation time are $N = 5 \cdot 10^3$ and $t = 5 \cdot 10^3$, respectively.

contribute a factor e^{-s} in the modified dynamics. The s-modified generator $\tilde{\mathbf{G}}$ is

$$\tilde{\mathbf{G}} = \begin{pmatrix} -\lambda & \mu & 0 & \dots & & \\ \lambda & -\lambda-\mu & \mu & & & \\ 0 & \lambda & -\lambda-\mu & & & \\ \vdots & & & \ddots & & \\ & & & & -\lambda-\mu & \mu \\ & & & & \lambda & -\lambda-\mu+\lambda\mathrm{e}^{-s} \end{pmatrix}. \tag{5.34}$$

This can be derived from a Master equation in the extended configuration–current space (a configuration here is given by the occupation number, while the number of packets lost represents a total current) as done, for semi-Markov processes, in section 4.3. The negative terms of $\{\tilde{\mathbf{G}}\}_{\mathsf{N}+1,\mathsf{N}+1}$ correspond to the "escape" rates from the state with occupation N. The positive entry $\lambda\mathrm{e}^{-s}$ encodes for the jumps that does not modify the state though increasing the count of the lost particles.

Figure 5.5 shows that the SCGF computed by means of the cloning method of the chapter 4 has an excellent agreement with the smallest eigenvalue of $\tilde{\mathbf{G}}$.

5.5 Discussion

While physicists have been regarding the large deviation theory as the way to formulate Statistical Mechanics, teletraffic engineers and operational researchers have been using large deviation results to estimate the likelihood that a demand in service overflows the available resources. A central role in teletraffic engineering is played by the effective bandwidth $\mathsf{e}(s,t)$, a function that provides a criterion to decide whether a given quality of service (QoS) can be maintained. This chapter reports some observations regarding the analogies between the effective bandwidth and the SCGF $e(s)$. It also suggests that the cloning methods of section 4 is a general scheme for the evaluation of $\mathsf{e}(s,t)$.

This is akin to the other problems discussed in the previous chapter; in fact, realistic queuing models must incorporate memory as they convey the patterns of human dynamics, which are non-Markovian [3, 12]. Despite the fact that the specialised literature is rich in exact solutions (mainly for Markov processes, see, e.g., [4, 14, 18, 24, 27] and the references therein), a systematic and general way to compute the effective bandwidth can be of interest.

It is important to remark that the effective bandwidth is a finite-time quantity and for short times the use of the cloning method is not justified (we recall that this method has been conceived to cope with the exponential expansion/contraction of the weight associated to each trajectory). Hence, it would be interesting to estimate the conditions under which the standard Monte Carlo calculations outperform the cloning method. We also mention that the long-time limit of the effective bandwidth is equally interesting, as demonstrated in Elwalid and Mitra [8] for Markov processes. This leaves ground for further development.

Bibliography

[1] D. Anick, D. Mitra, and M. M. Sondhi. Stochastic theory of a data-handling system with multiple sources. *Bell Syst. Tech. J.*, 61(8):1871–1894, 1982.

[2] A. Badescu, L. Breuer, A. Da Silva Soares, G. Latouche, M. Remiche, and D. A. Stanford. Risk processes analyzed as fluid queues. *Scand. Actuarial J.*, 2005(2): 127–141, 2005.

[3] A. L. Barabasi. The origin of bursts and heavy tails in human dynamics. *Nature*, 435(May), 2005.

[4] J. A. Bucklew. *Large Deviation Techniques in Decision, Simulation, and Estimation.* Wiley, New York NY. Wiley, 1990.

[5] M. R. Crivellari, A. Faggionato, and D. Gabrielli. Averaging and large deviation principles for fully-coupled piecewise deterministic markov processes and applications to molecular motors. *Markov Process. Relat.*, 16(3):497–548, 2010.

[6] M. H. A. Davis. Piecewise-deterministic Markov processes: A general class of non-diffusion stochastic models. *J. R. Stat. Soc. B*, pages 353–388, 1984.

[7] M. H. A. Davis. *Markov Models & Optimization*. Chapman & Hall, London UK. Taylor & Francis, 1993.

[8] A. I. Elwalid and D. Mitra. Effective bandwidth of general Markovian traffic sources and admission control of high speed networks. *IEEE/ACM Trans. Netw.*, I(3), 1993.

[9] A. Faggionato, D. Gabrielli, and M. R. Crivellari. Non-equilibrium thermodynamics of piecewise deterministic markov processes. *J. Stat. Phys.*, 137(2):259–304, 2009.

[10] W. Fischer and K. Meier-Hellstern. The Markov-modulated Poisson process (MMPP) cookbook. *Perform. Eval.*, 18(2):149–171, 1993.

[11] J. M. Harrison. *Brownian Models of Performance and Control*. Cambridge University Press, New York NY, 2013.

[12] B. A. Huberman. Social dilemmas and internet congestion. *Science*, 277(5325): 535–537, 1997.

[13] P. G. Hufton, Y. T. Lin, T. Galla, and A. J. McKane. Intrinsic noise in systems with switching environments. *Phys. Rev. E*, 93(5):052119, 2016.

[14] F. Kelly. Notes on effective bandwidths. In F. Kelly, S. Zachary, and I. Ziedins, editors, *Stochastic Networks: Theory and Application*, pages 141–168. Oxford University Press, Oxford UK, 1996.

[15] F. Kelly and E. Yudovina. *Stochastic Networks*. Cambridge University Press, Cambridge UK, 2014.

[16] G. Kesidis, J. Walrand, and C.-S. Chang. Effective bandwidths for multiclass Markov fluids and other ATM sources. *IEEE/ACM Trans. Netw.*, 1(4):424–428, 1993.

[17] V. Lecomte and J. Tailleur. A numerical approach to large deviations in continuous time. *J. Stat. Mech.*, 2007(03):P03004–P03004, 2007.

[18] J. Lewis and R. Russel. An introduction to large deviations for teletraffic engineers (DIAS Technical Report), 1996.

[19] Y. T. Lin and T. Galla. Bursting noise in gene expression dynamics: linking microscopic and mesoscopic models. *J. R. Soc. Interface*, 13(114):20150772, 2016.

[20] P. A. P. Moran. A Probability Theory of Dams and Storage Systems. *Australian J. of Applied Science*, 5:116–124, 1954.

[21] J. Realpe-Gomez, T. Galla, and A. J. McKane. Demographic noise and piecewise deterministic Markov processes. *Phys. Rev. E*, 86(1):011137, 2012.

[22] J. Realpe-Gomez, M. Baudena, T. Galla, A. J. McKane, and M. Rietkerk. Demographic noise and resilience in a semi-arid ecosystem model. *Ecol. Complexity*, 15: 97–108, 2013.

[23] J. W. Roberts. Performance evaluation and design of multiservice networks. Tech-

nical report, Office for Official Publications of the European Communities, Luxembourg, 1992.

[24] A. Shwartz and A. Weiss. *Large Deviations For Performance Analysis: Queues, Communication and Computing*. Chapman & Hall, London UK, 1995.

[25] D. A. Stanford, G. Latouche, D. G. Woolford, D. Boychuk, and A. Hunchak. Erlangized fluid queues with application to uncontrolled fire perimeter. *Stoch. Models*, 2007.

[26] W. J. Stewart. *Probability, Markov Chains, Queues, and Simulation: The Mathematical Basis of Performance Modeling*. Princeton University Press, Princenton NJ, 2009.

[27] A. Weiss. An introduction to large deviations for communication networks. *IEEE J. Sel. Areas Commun.*, 13(6):938–952, 1995.

[28] S. Zeiser, U. Franz, O. Wittich, and V. Liebscher. Simulation of genetic networks modelled by piecewise deterministic Markov processes. *IET Syst. Biol.*, 2(3):113–35, 2008.

[29] S. Zeiser, U. Franz, and V. Liebscher. Autocatalytic genetic networks modeled by piecewise-deterministic Markov processes. *J. Math. Biol.*, 60(2):207–246, 2010.

6 Conclusions and outlook

This thesis is concerned with the effects of memory in non-equilibrium systems. We considered models governed by stochastic dynamical rules, which can be encoded into a set of inter-event waiting-time distributions with transition events not satisfying the detailed balance. Such waiting-time distributions can describe well physical systems that are driven out of equilibrium (by, e.g., self-propulsion or the environment) or technological systems that follow man-made directives. More specifically, we analysed large deviations for these models. The results obtained are discussed in the last sections of each chapter. In this final chapter we summarise the results, add a few comments from a general point of view, and re-pose some questions left open throughout the thesis.

The first chapter is an original overview of the established mathematics of stochastic processes, where we introduced the dichotomies of memory-less/non-Markovian processes and equilibrium/non-equilibrium states, and the fuzzier distinction between typical and rare trajectories in large deviation theory. In the absence of a unified framework that comprehends all the non-equilibrium systems, out first step in chapter 2 was to study a specific driven model of particles on lattice, where we established new results, through both analysis and computer simulations. We studied the open-boundary on-off zero-range process (on-off ZRP), a model that incorporates memory by means of an additional "phase" variable (such phase is also referred to as a "clock"). This model is the open-boundary version of a non-Markovian ZRP introduced in Hirschberg et al. [10], where the temporal correlations derive from the following mechanisms: the particles are blocked on a lattice site ("phase off") when a new particle arrives (this facilitates congestion, i.e., the accumulation of particles on the site), the block being removed ("phase on") after an exponentially distributed waiting time with parameter c. The simple, but important, message of chapter 2 is that, at first sight, the effects of time correlations are hidden. In fact, the stationary-state solution of the one-site system (equation (2.18)) can be written as in the Markovian case (see equation (2.7)), with an effective on-site interaction $w_{c,n}$. This means that it is not possible to distinguish the on-off dynamics from a standard memory-less ZRP based only on the site-occupation distribution, hence we

cannot predict phase-dependent interactions when we are totally unaware of them. However, the presence of the on and off phases alters the statistics of the outwards particle hops. This becomes important in the spatially-extended system where each site receives particles, from its neighbours, according to a non-Markovian process. As a consequence, a product form solution is in general not expected and we have relied on a mean-field approach for the analytical treatment. This approximation consists of replacing the true arrival process on each site with a Poisson (i.e., memory-less) process, while keeping exact information about the on-site particle departure as well as the lattice topology. This procedure can be applied in principle to decouple non-Markovian ZRPs on an arbitrary lattice, provided that it is possible to solve the consistency equation for the fugacities (as done in section 2.5 for the one-dimensional chain). We found that, in the chain topology studied here, the mean-field approach is accurate for large values of c and gives an analytical approximate estimate c_{mf} for the congestion threshold. However, it fails when c is small, as the temporal correlations are more important in this case.

The memory effects at the fluctuating level appear more interesting even in the single-site case, as seen in chapter 3. The theory of large deviation allows us to quantify the change in the fluctuating behaviour due to the addition of temporal correlations. Fluctuations close to the mean current are obtained by analytic continuation of the stationary state and are indistinguishable from the fluctuations in the memory-less ZRP. However, under certain conditions, large current fluctuations are optimally realized by the instantaneous piling up of particles (a situation that is mathematically equivalent to the congestion transition) on site and the statistics of such fluctuations change abruptly. This behaviour is referred to as a dynamical phase transition. We treated separately the case where the particles are independent (if we exclude residual interaction due to the on-off dynamics) and the case where the particles have a weak attractive interaction (encoded in constant departure rates).

- In the absence of direct inter-particle interactions we have found analytically a memory-induced dynamical first-order phase transition, corresponding to a non-analyticity of the scaled cumulant generating function (SCGF) $e(s)$ at particular value s_1, which depends on c. When $s \leq s_1$ the fluctuations can be described using an s-dependent effective interaction factor $w_{c,n,s}$. A very interesting point to mention here is that this occurs only if the parameter c is smaller than a threshold value c_0; otherwise, current fluctuations are unaffected (in the infinite-time limit) by the on-off dynamics[1].

- The system with constant departure rates, i.e., attractive inter-particle interaction,

[1] This is valid only when a real c_0 exists, see equations (3.23) and (3.24).

undergoes second-order (at $s = s_2$) as well as first-order (at $s = s_1$) dynamical phase transitions. The state of the system during small fluctuation events has the same form as the stationary state, but with an s-dependent effective interaction factor $w_{c,s}$, only when $s_2 \leq s \leq s_1$. We proved that the exact phase boundaries and the large deviation function of this regime are encoded in the "reduced" operator $\tilde{H}^\star$ (defined in equation (3.18)), which has the same structure as the s-modified Hamiltonian of the standard ZRP, but with the s-dependent interactions $w_{c,s}$.

- Numerical tests confirm the presence of the predicted c-dependent dynamical phase transitions in both cases within numerical accuracy. The separation between a small-fluctuation regime, with a memory independent SCGF, and high-fluctuation regimes, where memory plays a more obvious role, is a feature also found numerically in the spatially-extended ZRP.

In summary, for the model explored in chapters 2 and 3, time correlations can be absorbed in an effective memory-less description for the steady state, but can emerge at the fluctuating level and alter the probability of observing rare phenomena. This may be connected to the behaviour of stochastic systems with interacting fast and slow degrees of freedom, as studied, e.g., in reference [2] or to the effect of bulk hidden nodes in biochemical networks [17]. It would be of interest to explore more details of the on-off ZRP and look for similar memory effects in other complex systems. Further aspects regarding the on-off ZRP that could be addressed in future research are:

- For the constant departure rate case, we have used the integral representation of the reduced operator $\tilde{H}^\star$ in order to find an approximate solution for the fluctuations corresponding to $s < s_2$ and $s > s_1$. It would be interesting to solve the eigenproblem (3.2) for the exact s-modified Hamiltonian $\tilde{H}$ of equation (3.1), which provides exact information about the strongly fluctuating regimes. This would probably be relevant for Queueing Theory: in fact, stochastic models with on-off dynamics are widely used in performance evaluation of queueing systems [15, 18].

- A further research direction is to look for a threshold value of c that separates a regime where the factorized solution is permitted from a regime where it is not; this is suggested by the presence of the threshold c_0 for the dynamical phase transition in the independent-particle case; in fact, in the absence of such dynamical-phase transitions, the current fluctuations are identical to the memory-less ZRP (with independent particles) which has a factorised steady state. At the present stage we do not have any proof that such a factorised solution exists for the on-off ZRP.

- It would also be interesting to explore the limit case $c \to 0^+$ for the TA model with

independendent particles. In this situation, in fact, the critical point s_1 approaches 0 and the mean current would have a first-order jump. This is reminiscent of the "glassy" behaviour described, e.g., in Garrahan et al. [5, 6].

The chapters 2 and 3 left interesting open questions about the predictive power of effective theories for real-world systems, where rare events can be of crucial importance. Hence a natural step forward is to find a way to probe *systematically* large deviations in more realistic models. We remark here that the analysis of non-Markovian models with hidden variables, such as the on-off ZRP, benefited of a Markovian representation in the joint configuration–phase space; this permitted us to derive both analytical and numerical results using some established tools for Markov processes. However, it is not hard to conceive non-Markovian models that are more appropriate to model real-world systems and cannot be represented in terms of hidden variables. In order to explore these models, we devised a general cloning algorithm and demonstrated it in chapter 4. This is another important result of this thesis, as it allows further numerical studies of large deviations in non-Markovian stochastic system, regardless of the mechanism that encodes for the memory. We mention that finite-ensemble errors for the cloning method in Markovian setting have been addressed in very recent literature [8, 9, 14] and a method with "feedback control" has been shown to mitigate such errors in Nemoto et al. [13]; similar results may be found in non-Markovian setting. In chapter 4 we tested the method on three selected non-Markovian models (namely, two simple semi-Markov models of ion channel and the history-dependent totally asymmetric exclusion process (TASEP) of Harris [7]) and showed that it seamlessly reproduces, within numerical accuracy, exact large deviation results obtainable by other means. It is worth mentioning that the cloning simulations for the history-dependent TASEP suggested the presence of a dynamical phase transition at a critical current value of j. Interestingly such a critical value is larger than the one obtained on the basis of an effective Markovian theory, while the typical behaviour of this model is only slightly modified by the presence of temporal correlations. This again seems to support the generic claim that the memory strongly affects the rare behaviour even when the typical behaviour is not noticeably altered.

In chapter 4 (and similarly in chapter 1), we reserved more attention to semi-Markov processes rather than to generic non-Markovian processes, as their long-time behaviour can be better understood in term of the generalised Master equation. Our opinion is that there still is ground for further development. We stress here that some important results (obtained in references [3, 16, 19]), show that, in order to have reversibility in a semi-Markov process, we need that both the number of transitions and the durations of the intervals between events in a trajectory and its reverse are "balanced" (see section 1.4.2). This suggests that in order to explore fluctuation symmetries in non-Markovian

systems, it is appropriate to consider functionals whose contributions take into account, simultaneously, configuration changes *and* inter-event times and hence define generic currents

$$\boxed{J[w(t)] = \sum_{i=0}^{n-1} \theta_{x_{i+1},x_i}(t_{i+1} - t_i),} \tag{6.1}$$

which generalise both the functionals of type A and B (defined in equations (1.111) and (1.113), respectively) more commonly found in scientific literature [6, 11]. To the best of our knowledge, non-equilibrium semi-Markov systems have been mainly studied by means of type A observables [1, 4, 12], while further progress could involve the statistics of the inter-event times.

Finally, in the second-to-last chapter, we reported some observations concerning analogies between large deviations in physics and teletraffic engineering, which complement the well-known analogies between queuing networks and interacting-particle systems (discussed at the beginning of chapter 1). Chapter 5 is intended to present a preliminary analysis on the applicability of the cloning method to the problem of performance evaluation in teletraffic engineering. This could feed further numerical studies.

In conclusion, a general understanding of non-equilibrium systems with memory remains a challenge. We wish to continue the work along the lines drawn in this thesis and contribute more to both the relevant applied and theoretical research.

Bibliography

[1] D. Andrieux and P. Gaspard. The fluctuation theorem for currents in semi-Markov processes. *J. Stat. Mech.*, 2008(11):P11007, 2008.

[2] S. Bo and A. Celani. Entropy production in stochastic systems with fast and slow time-scales. *J. Stat. Phys.*, 154(5):1325–1351, 2014.

[3] M. K. Chari. On reversible semi-Markov processes. *Oper. Res. Lett.*, 15(3):157–161, 1994.

[4] M. Esposito and K. Lindenberg. Continuous-time random walk for open systems: Fluctuation theorems and counting statistics. *Phys. Rev. E*, 77(5):051119, 2008.

[5] J. P. Garrahan, R. L. Jack, V. Lecomte, E. Pitard, K. van Duijvendijk, and F. van Wijland. Dynamical first-order phase transition in kinetically constrained models of glasses. *Phys. Rev. Lett.*, 98(19):195702, 2007.

[6] J. P. Garrahan, R. L. Jack, V. Lecomte, E. Pitard, K. van Duijvendijk, and F. van Wijland. First-order dynamical phase transition in models of glasses: an approach based on ensembles of histories. *J. Phys. A: Math. Gen.*, 42(7):075007, 2009.

[7] R. J. Harris. Fluctuations in interacting particle systems with memory. *J. Stat. Mech.*, 2015(7):P07021, 2015.

[8] E. G. Hidalgo and V. Lecomte. Discreteness effects in population dynamics. *J. Phys. A: Math. Gen.*, 49(20):205002, 2016.

[9] E. G. Hidalgo, T. Nemoto, and V. Lecomte. Finite-time and -size scalings in the evaluation of large deviation functions - Part II: Numerical approach in continuous time. 2016.

[10] O. Hirschberg, D. Mukamel, and G. M. Schütz. Condensation in temporally correlated zero-range dynamics. *Phys. Rev. Lett.*, 103(9):090602, 2009.

[11] R. L. Jack and P. Sollich. Effective interactions and large deviations in stochastic processes. *Eur. Phys. J. Spec. Top.*, 224(12):2351–2367, 2015.

[12] C. Maes, K. Netočný, and B. Wynants. Dynamical fluctuations for semi-Markov processes. *J. Phys. A: Math. Gen.*, 42(36):365002, 2009.

[13] T. Nemoto, F. Bouchet, R. L. Jack, and V. Lecomte. Population-dynamics method with a multicanonical feedback control. *Phys. Rev. E*, 93(6):062123, 2016.

[14] T. Nemoto, E. G. Hidalgo, and V. Lecomte. Finite-time and -size scalings in the evaluation of large deviation functions - Part I: Analytical study using a birth-death process. 2016.

[15] M. F. Neuts. *Matrix-geometric Solutions in Stochastic Models: An Algorithmic Approach*. Dover Publications, New York NY, 1981.

[16] H. Qian and H. Wang. Continuous time random walks in closed and open single-molecule systems with microscopic reversibility. *EPL*, 76(1):15–21, 2007.

[17] K. J. Rubin, K. Lawler, P. Sollich, and T. Ng. Memory effects in biochemical networks as the natural counterpart of extrinsic noise. *J. Theo. Biol.*, 357:245–267, 2014.

[18] W. J. Stewart. *Probability, Markov Chains, Queues, and Simulation: The Mathematical Basis of Performance Modeling*. Princeton University Press, Princenton NJ, 2009.

[19] H. Wang and H. Qian. On detailed balance and reversibility of semi-Markov processes and single-molecule enzyme kinetics. *J. Math. Phys.*, 48(1):1–15, 2007.

A | Spectrum and integral representation

In this appendix, we detail the procedure to diagonalise the infinite-dimensional tridiagonal matrix

$$\mathbb{H} = \begin{pmatrix} -b_0 & c & 0 & 0 & \dots \\ a & -b & c & 0 & \\ 0 & a & -b & c & \\ \vdots & & & & \ddots \end{pmatrix}. \tag{A.1}$$

This has been used in chapter 3 to derive the integral representation (3.37) for the special tridiagonal operator $\tilde{H}^\star$ of equation (3.18). Such a representation is related to the one of Karlin and McGregor [2] for the transition probabilities (1.25) (see also refence [1]) First, we transform $\mathbb{H}$ into the symmetric form

$$\mathbb{H}' := \Psi \mathbb{H} \Psi^{-1} = \begin{pmatrix} -b_0 & \sqrt{ac} & 0 & 0 & \dots \\ \sqrt{ac} & -b & \sqrt{ac} & 0 & \\ 0 & \sqrt{ac} & -b & \sqrt{ac} & \\ \vdots & & & & \ddots \end{pmatrix}, \tag{A.2}$$

where

$$\Psi := \begin{pmatrix} \sqrt{c/a} & 0 & 0 & 0 & \dots \\ 0 & \sqrt{c/a}^2 & 0 & 0 & \\ 0 & 0 & \sqrt{c/a}^3 & 0 & \\ \vdots & & & & \ddots \end{pmatrix}, \tag{A.3}$$

and consider the following eigenproblem for $\mathbb{H}'$

$$\mathbb{H}'|\psi'\rangle = \lambda|\psi'\rangle, \tag{A.4}$$

where the eigenvector $|\psi'\rangle$ has components ψ'_n, $n = 0, 1, 2, \dots$.

A.1 Spectrum of $\mathbb{H}$

Equation (A.2) corresponds to the generator of an unbounded bidimensional random walk at equilibrium. For $n > 0$, the generic row equation of A.4

$$\sqrt{ac}\,\psi_{n-1} - (b+\lambda)\psi_n + \sqrt{ac}\,\psi_{n+1} = 0 \tag{A.5}$$

implies, after a bilateral Z-transform $\tilde{\psi}(z) = \sum_{n=-\infty}^{+\infty} z^n \psi_n$,

$$\lambda = -b + \sqrt{ac}(z + z^{-1}). \tag{A.6}$$

To guarantee the convergence of $\tilde{\psi}(z)$ we choose the parametrization $z(k) = \mathrm{e}^{ik}$, where $k \in (0, \pi]$, and write

$$\boxed{\lambda(k) = -b + 2\sqrt{ac}\cos(k).} \tag{A.7}$$

The associated eigenvector has components $\sqrt{\frac{2}{\pi}}\sin(nk+\varphi)$. Imposing this on the first row equation, we get

$$[-b_0 - \lambda(k)]\left(\mathrm{e}^{i\varphi} - \mathrm{e}^{-i\varphi}\right) + \sqrt{ac}\left(\mathrm{e}^{i\varphi}\mathrm{e}^{ik} - \mathrm{e}^{-i\varphi}\mathrm{e}^{-ik}\right) = 0 \tag{A.8}$$

which gives the following expressions

$$\mathrm{e}^{i2\varphi} = \frac{-b_0 - \lambda(k) + \mathrm{e}^{-ik}\sqrt{ac}}{-b_0 - \lambda(k) + \mathrm{e}^{ik}\sqrt{ac}} = \frac{-b_0 + b - \sqrt{ac}\mathrm{e}^{ik}}{-b_0 + b - \sqrt{ac}\mathrm{e}^{-ik}} = \frac{y\mathrm{e}^{ik} - 1}{y\mathrm{e}^{-ik} - 1}, \tag{A.9}$$

where

$$y = \sqrt{ac}/(b - b_0). \tag{A.10}$$

This defines a continuous spectrum with associated non-normalisable eigenvector

$$\boxed{|\psi'(k)\rangle = \sum_{n=0}^{\infty} \sqrt{\frac{2}{\pi}}\sin(nk+\varphi)|e_n\rangle.} \tag{A.11}$$

We now look for a normalisable solution with geometric ansatz $|\psi'\rangle = (1, \mathsf{y}, \mathsf{y}^2, \mathsf{y}^3, \ldots)/C$. The first row equation of the eigenproblem (A.4) reads

$$-(b_0 + \lambda) + \sqrt{ac}\,\mathsf{y} = 0, \tag{A.12}$$

that is

$$\mathsf{y} = \frac{(b_0 + \lambda)}{\sqrt{ac}}, \tag{A.13}$$

while

$$\sqrt{ac} - (b+\lambda)\,\mathsf{y} + \sqrt{ac}\,\mathsf{y}^2 = 0 \tag{A.14}$$

implies

$$\mathsf{y} = \frac{(b+\lambda) \pm \sqrt{(b+\lambda)^2 - 4ac}}{2\sqrt{ac}}. \tag{A.15}$$

Combining (A.13) and (A.15), we get an expression for the eigenvalue corresponding to $|\psi'\rangle$

$$\boxed{\lambda = \frac{ac}{b - b_0} - b_0,} \tag{A.16}$$

which yields the explicit expression $\mathsf{y} = y$. Now, the normalisation condition $\langle\psi'|\psi'\rangle = 1$, where $\langle\psi'|$ is the left eigenvalue of $\mathbb{H}'$ and has the same components as $|\psi'\rangle$ ($\mathbb{H}'$ being real symmetric) demands that $y < 1$ and $C^2 = (1 - y^2)$, hence

$$\boxed{|\psi'\rangle = \sqrt{1-y^2} \sum_{n=0}^{\infty} y^n |e_n\rangle.} \tag{A.17}$$

Notice that $\lambda \neq \lambda(k)$.

To ensure that equations (A.11) and (A.17) form a complete basis, we need to prove that

$$\begin{cases} \int_0^\pi |\psi'(k)\rangle\langle\psi'(k)|\mathrm{d}k = \mathbb{1} & \text{if } y \geq 1, \\ \int_0^\pi |\psi'(k)\rangle\langle\psi'(k)|\mathrm{d}k + |\psi'\rangle\langle\psi'| = \mathbb{1} & \text{otherwise.} \end{cases} \tag{A.18}$$

To this end, we evaluate the integral

$$\begin{aligned} \int_0^\pi \langle n|\psi'(k)\rangle\langle\psi'(k)|m\rangle\,\mathrm{d}k &= \frac{2}{\pi}\int_0^\pi \sin(km+\varphi)\sin(kn+\varphi)\,\mathrm{d}k \\ &= \frac{2}{\pi}\int_0^\pi \frac{1}{4}\left(-\mathrm{e}^{-ik(m+n)}\mathrm{e}^{-2i\varphi} - \mathrm{e}^{ik(m+n)}\mathrm{e}^{2i\varphi} + \mathrm{e}^{ik(m-n)} + \mathrm{e}^{-ik(m-n)}\right)\mathrm{d}k. \end{aligned} \tag{A.19}$$

From this we isolate the term that does not depend on φ, which yields

$$\frac{2}{\pi}\int_0^\pi \frac{1}{2}\cos(k(m-n))\,\mathrm{d}k = \frac{\sin(\pi(m-n))}{\pi(m-n)} = \delta_{n,m}. \tag{A.20}$$

Inserting equation (A.9) in the remaining term, we get

$$-\frac{1}{4\pi}\int_0^{2\pi} \frac{y^2\mathrm{e}^{-ik(n+2)} + y^2\mathrm{e}^{ik(n+2)} - 2y\mathrm{e}^{-ik(n+1)} - 2y\mathrm{e}^{ik(n+1)} + \mathrm{e}^{-ikn} + \mathrm{e}^{ikn}}{(y - \mathrm{e}^{-ik})(y - \mathrm{e}^{ik})}\,\mathrm{d}k, \tag{A.21}$$

where we also changed the integration boundaries using the fact that the integrand is symmetric with respect to $k = 0$. We now make the coordinate change $z = \mathrm{e}^{ik}$, define

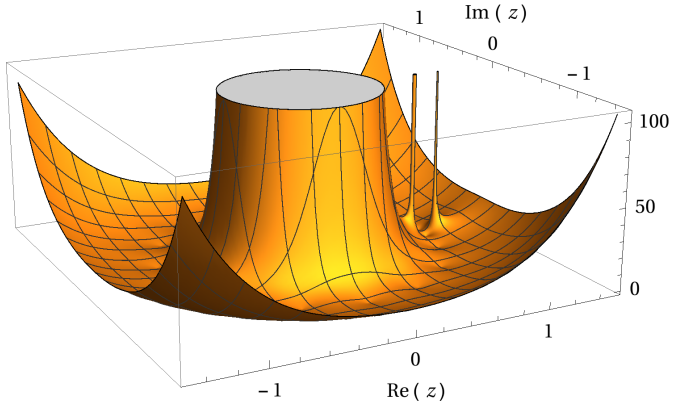

Figure A.1: Modulus of the integrand in equation (A.22) for $l = 6$ and $y = 1.1$. The poles are on the real axis, at $z = y$, $z = 1/y$, and $z = 0$. The integration countour $|z| = 1$ can engulf one or more of the poles in the real axis.

$l = m + n$ and get

$$-\frac{1}{4\pi i}\oint_{|z|=1}\frac{y^2z^{2l+4} - 2yz^{2l+3} + z^{2l+2} + y^2 - 2yz + z^2}{z^{l+2}(y-z)(y-z^{-1})}z\,\mathrm{d}z, \tag{A.22}$$

where $|z| = 1$ is the unit circle centred at $z = 0$. Its integrand is denoted by $f(z)$ and has singularities at $z = y$, $z = 1/y$, and $z = 0$, as shown in figure A.1 and can be solved using the residue theorem. The pole at $z = 0$ is always inside the integration contour and contributes a term

$$2\pi i\frac{1}{(l+2-1)!}\lim_{z\to 0}\frac{\mathrm{d}^{l+2-1}}{\mathrm{d}z^{l+2-1}}z^l f(z) = 2\pi i\,(1-y^2)y^l. \tag{A.23}$$

When $y > 1$ the pole at $z = 1/y$ contributes a second term that added to the previous one yields zero:

$$2\pi i\lim_{z\to 1/y}(z - 1/y)f(z) = -2\pi i\,(1-y^2)y^l. \tag{A.24}$$

When $y < 1$, the pole contributions at $z = y$ is

$$2\pi i\lim_{z\to y}(z-y)f(z) = 2\pi i\,(1-y^2)y^l, \tag{A.25}$$

hence the expression (A.22) is equal to $-(1-y^2)y^{m+n}$. This indeed cancels with the component $\langle e_n|\psi'\rangle\langle\psi'|e_m\rangle = (1-y^2)y^{n+m}$ in equation (A.18). For $y = 1$ the two cases are equivalent. In all the cases, the only remaining contribution is equation (A.20).

Finally, the left and right eigenvectors of $\mathbb{H}$ can be obtained as ([3])

$$|\psi(k)\rangle = \Psi^{-1}|\psi'(k)\rangle, \qquad \langle\psi(k)| = \langle\psi'(k)|\Psi, \tag{A.26}$$

$$|\psi\rangle = \Psi^{-1}|\psi'\rangle, \qquad \langle\psi| = \langle\psi'|\Psi. \tag{A.27}$$

A.2 Spectrum of $\tilde{H}^\star$

If we replace in matrix (A.1) the entries a, b, b_0, and c with $\alpha+\delta e^s$, $\alpha+\delta+(\beta+\gamma)w_{c,s}$, $\alpha+\delta$, and $(\beta e^{-s}+\gamma)w_{c,s}$ respectively, we obtain the special operator $\tilde{H}^\star$ seen in section 3.1.1.1 (equation (3.18)). With such replacements, the continuous spectrum equation (A.7) becomes equation (3.40) (with opposite sign, as we used there the quantum Hamiltonian formalism), while the leading eigenvalue (A.16) becomes $-A_0$ (defined in equation (3.12)). Equation (A.15) yields equation (3.39). Also, it is convenient to use a shorten notation, defining

$$\phi := \sqrt{c/a}, \tag{A.28}$$

which gives back equation (3.38).

A.3 Integral representation of the generating function

We now use the spectral properties found in the previous sections on the generating function $\langle 1|e^{-\tilde{H}^\star t}|\tilde{P}(0)\rangle$. Let us impose a initial condition of Boltzmann-type for the $|\tilde{P}(0)\rangle$, namely the geometric distribution with parameter x of equation (3.34). Then, the generating function of the total current at time t is

$$\langle 1|e^{-\tilde{H}^\star t}|\tilde{P}(0)\rangle = (1-x)\sum_{n,m=0}^{\infty} x^n \langle e_m|e^{-\tilde{H}^*}t|e_n\rangle, \tag{A.29}$$

where $\langle e_m|$ ($|e_n\rangle$) is a row (column) vector with a '1' in the m (n) position and '0' elsewhere. The vectors of equations (A.26) and (A.27) form a complete set, i.e.,

$$\int_0^\pi |\psi(k)\rangle\langle\psi(k)|\mathrm{d}k + |\psi\rangle\langle\psi| = \mathbb{1}. \tag{A.30}$$

Inserting this representation of the identity in equation (A.29), the r.h.s. become

$$(1-x)\sum_{n,m=0}^{\infty} x^n \int_0^{\pi} \langle e_m|\psi(k)\rangle\langle\psi(k)|e_n\rangle \mathrm{e}^{-\epsilon(k)t}\,\mathrm{d}k \\ + \Theta(1-y)(1-x)\sum_{m,n=0}^{\infty} x^n\phi^{n-m}(1-y^2)y^{n+m}\mathrm{e}^{-A_0 t}, \quad \text{(A.31)}$$

where Θ denotes the Heaviside step function. Using the fact the eigenvectors have period 2π and are odd in k, the expression (A.31) can be rewritten as

$$\frac{1}{2\pi}\phi^{n-m}\int_0^{2\pi}\left(\mathrm{e}^{ik(n-m)} - \mathrm{e}^{ik(n+m)}\mathrm{e}^{i2\varphi}\right)\mathrm{e}^{-\epsilon(k)t}\,\mathrm{d}k \\ + \Theta(1-y)(1-x)\sum_{m,n=0}^{\infty} x^n\phi^{n-m}(1-y^2)y^{n+m}\mathrm{e}^{-A_0 t}. \quad \text{(A.32)}$$

Using equation (3.38), it becomes

$$\frac{1}{2\pi}\phi^{n-m}\oint_{|\zeta|=1}\left(\zeta^{n-m} - \zeta^{n+m}\frac{1-\zeta y}{\zeta - y}\right)\mathrm{e}^{-\varepsilon(\zeta)t}\,\mathrm{d}\zeta \\ + \Theta(1-y)(1-x)\sum_{m,n=0}^{\infty} x^n\phi^{n-m}(1-y^2)y^{n+m}e^{-A_0 t}, \quad \text{(A.33)}$$

where $\zeta = \mathrm{e}^{ik}$ and $\varepsilon(\zeta) = \epsilon[k(\zeta)]$. Deforming the integration contour to C_1 for the first term in the integrand and to C_2 for the second term, we obtain the representation (3.37). The last term in expression (A.33) cancels out with a pole contribution in $\zeta = y$ for $y < 1$.

Bibliography

[1] N. G. Van Kampen. *Stochastic processes in physics and chemistry*. Elsevier, Amsterdam The Netherlands, Oxford UK, third edition, 1992.

[2] S. Karlin and J. L. McGregor. The differential equations of birth-and-death processes, and the Stieltjes moment problem. *T. Am. Math. Soc.*, 85(2):489–489, 1957.

[3] A. Rákos and R. J. Harris. On the range of validity of the fluctuation theorem for stochastic Markovian dynamics. *J. Stat. Mech.*, 2008(05):P05005, 2008.

B | s-modified effective Hamilitonian

In this appendix we report the explicit eigenproblem for the s-modified "collapsed" probabilities. Similarly to what we have seen for the stationary state, it is possible to define an s-modified Hamiltonian for the components $\tilde{P}_n(t) = \tilde{P}_{n,\text{ON}}(t) + \tilde{P}_{n,\text{OFF}}(t)$ of the vectors in the reduced state space. These satisfy

$$\frac{\mathrm{d}}{\mathrm{d}t}\tilde{P}_0(t) = \mu_1(\gamma + \beta \mathrm{e}^{-s})\tilde{P}_{1,\text{ON}}(t) - (\alpha + \delta)\tilde{P}_{0,\text{OFF}}(t), \tag{B.1}$$

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\tilde{P}_n(t) =& \mu_{n+1}(\gamma + \beta \mathrm{e}^{-s})\tilde{P}_{n+1,\text{ON}}(t) + (\alpha + \delta \mathrm{e}^{s})\tilde{P}_{n-1}(t) \\ &- (\alpha + \delta)\tilde{P}_{n,\text{OFF}}(t) - \mu_n(\gamma + \beta)\tilde{P}_{n,\text{ON}}(t),\end{aligned} \tag{B.2}$$

where $n = 1, 2, \ldots$. Using $|\tilde{P}^\star\rangle$ for the right eigenvector with components $\tilde{P}^\star_0, \tilde{P}^\star_1, \tilde{P}^\star_2, \ldots$, the eigenproblem associated to the dynamics (B.1) and (B.2) can written as

$$\mu_1(\gamma + \beta \mathrm{e}^{-s})\tilde{P}^\star_{1,\text{ON}} - (\alpha + \delta)\tilde{P}^\star_{0,\text{OFF}} = A\tilde{P}^\star_0, \tag{B.3}$$

$$\begin{aligned}\mu_{n+1}(\gamma + \beta \mathrm{e}^{-s})\tilde{P}^\star_{n+1,\text{ON}} + (\alpha + \delta \mathrm{e}^{s})\tilde{P}^\star \\ - (\alpha + \delta)\tilde{P}^\star_{n,\text{OFF}} - \mu_n(\gamma + \beta)\tilde{P}^\star_{n,\text{ON}} = A\tilde{P}^\star_n,\end{aligned} \tag{B.4}$$

with $n = 1, 2, \ldots$. Let us assume that we are only aware of the structure of the stationary state, without realising the presence of two separated stages (viz., on and off). Then, we would be satisfied with using equations (3.3)-(3.5) with ansatz (3.15)-(3.16) on (B.4) and (B.3), to get

$$\mu_1(\gamma + \beta \mathrm{e}^{-s})p_{\text{ON},1,s}\tilde{P}^\star_1 - (\alpha + \delta)(1 - p_{\text{ON},0,s})\tilde{P}^\star_0 = A\tilde{P}^\star_0, \tag{B.5}$$

$$\begin{aligned}\mu_{n+1}(\gamma + \beta \mathrm{e}^{-s})p_{\text{ON},n,s}\tilde{P}^\star_{n+1} + (\alpha + \delta \mathrm{e}^{s})\tilde{P}^\star_{n-1} \\ - (\alpha + \delta)(1 - p_{\text{ON},n,s})\tilde{P}^\star_n - \mu_n(\gamma + \beta)p_{\text{ON},n,s}\tilde{P}^\star_n = A\tilde{P}^\star_n.\end{aligned} \tag{B.6}$$

In the quantum Hamiltonian formalism this, in fact, is the eigenproblem $\tilde{H}^\star|\tilde{P}^\star\rangle = A|\tilde{P}^\star\rangle$ with s-modified Hamiltonian

$$\tilde{H}^\star = \alpha(a^+ - 1) + \delta(\mathrm{e}^s a^+ - 1) + \gamma(a_s^{\star -} - d_s^\star) + \beta(\mathrm{e}^{-s} a_s^{*-} - d_s^*), \tag{B.7}$$

where

$$\tilde{a}^{\star-} = \begin{pmatrix} 0 & w_{1,c,s} & 0 & 0 & \dots \\ 0 & 0 & w_{2,c,s} & 0 & \\ 0 & 0 & 0 & w_{3,c,s} & \\ 0 & 0 & 0 & 0 & \\ \vdots & & & & \ddots \end{pmatrix}, \tag{B.8}$$

$$a^{+} = \begin{pmatrix} 0 & 0 & 0 & \dots \\ 1 & 0 & 0 & \\ 0 & 1 & 0 & \\ \vdots & & & \ddots \end{pmatrix}, \tag{B.9}$$

$d^{\star}_{sij} = \delta_{ij} w_{i,c,s}$ and $w_{i,c,s} = \mu_i p_{\mathrm{ON},i,s}$. Notice the $w_{i,c,s}$ is s-dependent. Since the eigenvector satisfies

$$a^{+}|\tilde{P}^{\star}\rangle = z_s^{-1} d^{\star}|\tilde{P}^{\star}\rangle, \tag{B.10}$$

$$a^{\star-}|\tilde{P}^{\star}\rangle = z_s|\tilde{P}^{\star}\rangle, \tag{B.11}$$

then the eigenproblem can be rewritten as

$$\alpha(z_s^{-1} d_s^{\star} - 1) + \delta(\mathrm{e}^{s} z_s^{-1} d_s^{\star} - 1) + \gamma(z_s - d_s^{\star}) + \beta(\mathrm{e}^{-s} z_s - d_s^{*})|\tilde{P}^{\star}\rangle = A|\tilde{P}^{\star}\rangle, \tag{B.12}$$

where the operator at the l.h.s. is diagonal and implies the form (3.12) for the eigenvalue, i.e., $A = A_0$, and the expression (3.13) for z_s.

As explained in the main text, the s-modified operator (B.7) is not a genuine Hamiltonian for the on-off ZRP dynamics. It can be seen as a biased generator that shares the small-fluctuation regime (including the leading eigenvalue A_0 and the critical points s_1 and s_2, which can be found as the values of s such that $\langle \tilde{P}^{\star}|\tilde{P}^{\star}\rangle$ and $\langle 1|\tilde{P}^{\star}\rangle$ diverge, respectively) with the original process. However, it does not predict the correct statistics of rare currents (corresponding to $s > s_1$ and $s < s_2$). In other words, the current of the process generated by $\tilde{H}^{\star}$ is, in the long time limit, identical to the original on-off ZRP for $s_2 \leq s \leq s_1$, while rare currents are optimally realised with different probabilities and without the involvement of the on-off mechanism.

C | Left eigenvector of the s-Hamiltonian

In this appendix we derive the left eigenvector $\langle\tilde{P}|$ of the s-Hamiltonian of equation (3.1) when $\mu_n = \mu$, $n > 0$. Assuming that its components satisfy

$$\tilde{P}^{\text{left}}_{n,\text{ON}} = p^{\text{left}}_{\text{ON},s}\tilde{P}^{\text{left}}_{n}, \tag{C.1}$$

$$\tilde{P}^{\text{left}}_{n,\text{OFF}} = (1 - p^{\text{left}}_{\text{ON},s})\tilde{P}^{\text{left}}_{n}, \tag{C.2}$$

$$\tilde{P}^{\text{left}}_{n+1} = \rho^{\text{left}}_{s}\tilde{P}^{\text{left}}_{n}, \tag{C.3}$$

we get, for $\langle\tilde{P}|\tilde{H} = A\langle\tilde{P}|$, the explicit column equations

$$-(1 - p^{\text{left}}_{\text{ON},s,0})(\alpha - A + c + \delta) + cp^{\text{left}}_{\text{ON},s,0} + \rho^{\text{left}}_{s}(1 - p^{\text{left}}_{\text{ON},s})\,(\alpha + \delta\mathrm{e}^{s}) = 0, \tag{C.4}$$

$$\rho^{\text{left}}_{s}(1 - p^{\text{left}}_{\text{ON},s})\,(\alpha + \delta\mathrm{e}^{s}) - p^{\text{left}}_{\text{ON},s,0}(\alpha - A + \delta) = 0, \tag{C.5}$$

$$-(1 - p^{\text{left}}_{\text{ON},s})(\alpha - A + c + \delta) + c\left(1 - p^{\text{left}}_{\text{ON},s,0}\right) + \rho^{\text{left}}_{s}(1 - p^{\text{left}}_{\text{ON},s})\,(\alpha + \delta\mathrm{e}^{s}) = 0, \tag{C.6}$$

$$\mu p^{\text{left}}_{\text{ON},s,0}\left(\gamma + \beta\mathrm{e}^{-s}\right) - \rho^{\text{left}}_{s}p^{\text{left}}_{\text{ON},s}(\alpha - A + \mu(\beta + \gamma) + \delta) + \left(\rho^{\text{left}}_{s}\right)^{2}(1 - p^{\text{left}}_{\text{ON},s})\,(\alpha + \delta\mathrm{e}^{s}) = 0, \tag{C.7}$$

$$\mu p^{\text{left}}_{\text{ON},s}\left(\gamma + \beta\mathrm{e}^{-s}\right) - \rho^{\text{left}}_{s}p^{\text{left}}_{\text{ON},s}(\alpha - A + \mu(\beta + \gamma) + \delta) + \left(\rho^{\text{left}}_{s}\right)^{2}(1 - p^{\text{left}}_{\text{ON},s})\,(\alpha + \delta\mathrm{e}^{s}) = 0, \tag{C.8}$$

where the factor $p^{\text{left}}_{\text{ON},s,0}$ is assumed to be different from $p^{\text{left}}_{\text{ON},s}$, in analogy with the right eigenproblem. Equations (C.4) and (C.5) give

$$-(1 - p^{\text{left}}_{\text{ON},s,0})(\alpha - A + c + \delta) + p^{\text{left}}_{\text{ON},s,0}(\alpha - A + \delta) + cp^{\text{left}}_{\text{ON},s,0} = 0, \tag{C.9}$$

which is verified for

$$p^{\text{left}}_{\text{ON},s,0} = 1/2, \tag{C.10}$$

while equations (C.7) and (C.8) imply

$$p^{\text{left}}_{\text{ON},s,0} = p^{\text{left}}_{\text{ON},s}. \tag{C.11}$$

After the substitution, the remaining equations are solved for $A = A_0$ and

$$\rho^{\text{left}}_{s} = (\gamma + \beta\mathrm{e}^{-s})/(\beta + \gamma). \tag{C.12}$$

With those constants, it is easy to verify that the ansatz (C.1)-(C.3) is consistent even in the general departure rate case. In fact, after substitution, all the terms containing μ_n cancel out.

Similarly to the column-vector space case of appendix B, we show that it is possible to work in a reduced space also with row-vectors and recover an eigenproblem

$$\langle \tilde{P}^\star | \tilde{H}^\star_{\text{left}} = A_0 \langle \tilde{P}^\star |, \tag{C.13}$$

where $\langle \tilde{P}^\star |$ has components $\tilde{P}^{\text{left}}_0, \tilde{P}^{\text{left}}_1, \tilde{P}^{\text{left}}_2, \ldots$ given by equation C.3 and $\tilde{H}^\star_{\text{left}}$ is defined below. In fact, summing pairwise the column equations of the complete left eigenproblem for $\tilde{H}$, we get

$$-\frac{1}{2}(\alpha+\delta-A)\tilde{P}^{\text{left}}_0 + (\alpha+\delta \mathrm{e}^s)\tilde{P}^{\text{left}}_1 = 0, \tag{C.14}$$

$$\frac{1}{2}\mu_n \tilde{P}^{\text{left}}_{n-1} - (\alpha+\delta+\frac{1}{2}\mu_n(\beta+\gamma)-A)\tilde{P}^{\text{left}}_n + (\alpha+\delta\mathrm{e}^s)\tilde{P}^{\text{left}}_{n+1} = 0, \tag{C.15}$$

$n = 1, 2, 3, \ldots$, which can be arranged as equation C.13 with $A = A_0$ and

$$\tilde{H}^\star_{\text{left}} = -\alpha(a^+ - 1) - \beta(\mathrm{e}^{-s} a^{\star-}_{\text{left}} - d^\star_{\text{left}}) - \gamma(a^{\star-}_{\text{left}} - d^\star_{\text{left}}) - \delta(\mathrm{e}^s a^+ - 1), \tag{C.16}$$

where $d^{\star-}_{\text{left}\,ij} = \frac{1}{2}\mu_i \delta_{ij}$ and

$$\alpha^{\star-}_{\text{left}} = \begin{pmatrix} 0 & \frac{1}{2}\mu_1 & 0 & 0 & \cdots \\ 0 & 0 & \frac{1}{2}\mu_2 & 0 & \\ 0 & 0 & 0 & \frac{1}{2}\mu_3 & \\ 0 & 0 & 0 & 0 & \\ \vdots & & & & \ddots \end{pmatrix}. \tag{C.17}$$

In this worth noting that $\tilde{H}^\star_{\text{left}} \neq \tilde{H}^\star$ but, as the terms containing μ_n cancel out with the solution (C.12) and (C.10), we also have $\langle \tilde{P}^\star | \tilde{H}^\star = A_0 \langle \tilde{P}^\star |$.

List of Abbreviations

ASEP	Asymmetric exclusion process
CRTW	Continuous-time random walk
DTI	Direction–time independence
GME	Generalised Master equation
LF transform	Legendre–Fenchel transform
l.h.s.	Left hand side
NESS	Non-equilibrium stationary state
QoS	Quality of Service
r.h.s.	Right hand side
SCGF	Scaled cumulant generating function
WTD	Waiting-time distribution
ZRP	Zero-range process